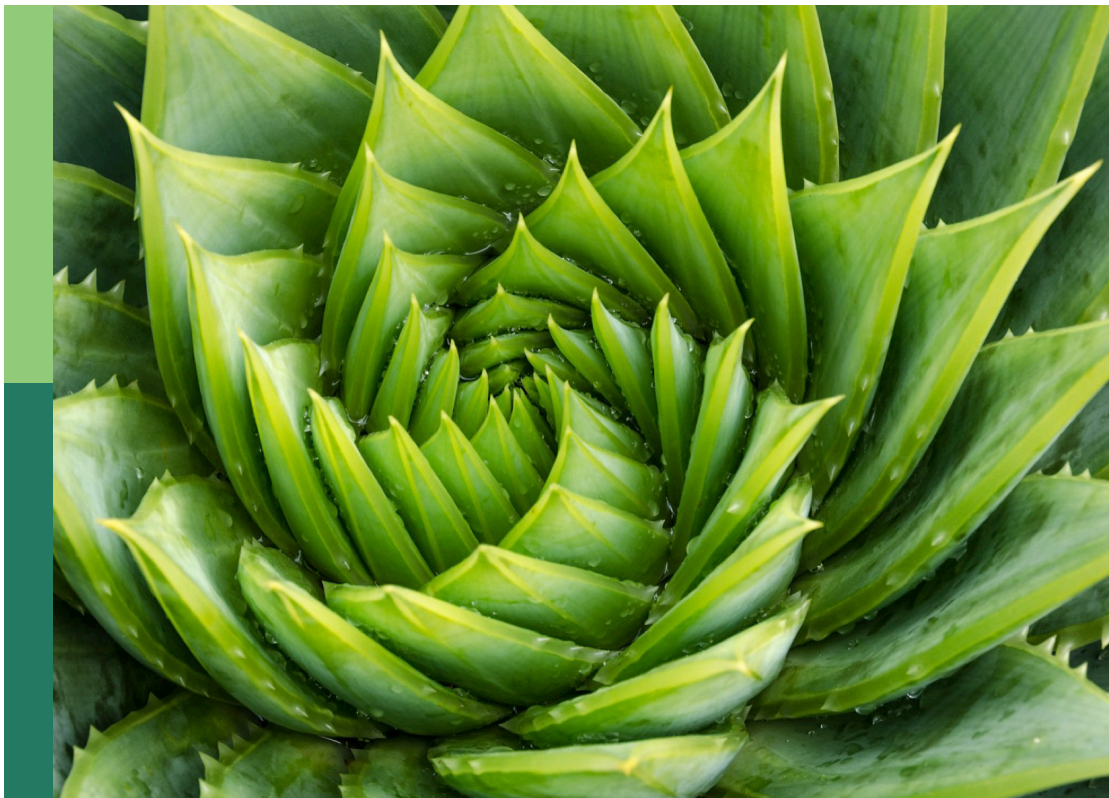
frontiers | Research Topics

The brassicaceae — agri-horticultural and environmental perspectives, volume II

Edited by
Naser A. Anjum, Om Parkash Dhankher, Narendra Tuteja, Sarvajeet Singh Gill and Juan F. Jimenez

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83251-563-1
DOI 10.3389/978-2-83251-563-1

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

The brassicaceae — agri-horticultural and environmental perspectives, volume II

Topic editors

Naser A. Anjum — Aligarh Muslim University, India

Om Parkash Dhankher — University of Massachusetts Amherst, United States

Narendra Tuteja — International Centre for Genetic Engineering and Biotechnology, India

Sarvajeet Singh Gill — Maharshi Dayanand University, India

Juan F. Jimenez — Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), Mexico

Citation

Anjum, N. A., Dhankher, O. P., Tuteja, N., Gill, S. S., Jimenez, J. F., eds. (2023). *The brassicaceae — agri-horticultural and environmental perspectives, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-563-1

Table of contents

Brassica napus

Brassica rapa

Brassica juncea

Brassica oleracea

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 20 December 2022
DOI 10.3389/fpls.2022.1108246

OPEN ACCESS

EDITED AND REVIEWED BY
Diego Rubiales,
Institute for Sustainable, Spain

*CORRESPONDENCE
Naser A. Anjum
naanjum1@myamu.ac.in

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

RECEIVED 25 November 2022
ACCEPTED 01 December 2022
PUBLISHED 20 December 2022

CITATION
Anjum NA, Gill SS, Dhankher OP, Tuteja N and Jimenez JF (2022) Editorial: The Brassicaceae — agri-horticultural and environmental perspectives, volume II. *Front. Plant Sci.* 13:1108246. doi: 10.3389/fpls.2022.1108246

Editorial: The Brassicaceae — agri-horticultural and environmental perspectives, volume II

Naser A. Anjum[1]*, Sarvajeet Singh Gill[2], Om P. Dhankher[3], Narendra Tuteja[4] and Juan F. Jimenez[5]

[1]Department of Botany, Aligarh Muslim University, Aligarh, India, [2]Stress Physiology and Molecular Biology Lab, Centre for Biotechnology, Maharshi Dayanand (MD) University, Rohtak, India, [3]Stockbridge School of Agriculture, University of Massachusetts Amherst, Amherst, MA, United States, [4]Plant Molecular Biology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India, [5]Institute for Scientific and Technological Research, San Luis Potosí, Mexico

KEYWORDS

brassicaceae family, agricultural perspective, horticultural perspective, environmental health, plant breeding, crop improvement

Editorial on the Research Topic
The Brassicaceae — agri-horticultural and environmental perspectives, volume II

Brassica is an old world genus and a monophyletic group within the Brassicaceae. Comprising about 35 species, the genus Brassica is represented by six interrelated cultivated brassicas, *Brassica napus*, *B. rapa*, *B. juncea*, *B. oleracea*, *B. nigra*, and *B. carinata*. Among the 18 research articles published in this volume II of the Research Topic '*The Brassicaceae - Agri-Horticultural and Environmental Perspectives*', *B. napus*, *B. rapa*, *B. juncea*, and *B. oleracea* were researched in 11, 4, 2, and 1 articles on these respective genus.

This second volume provides insights into the molecular-genetic nature of the least and/or unexplored aspects in the major species of genus Brassica to date. It includes: (i) the quantitative trait loci related to silique-associated traits (Zhao et al.); (ii) heterosis of yield-influencing traits (Aakanksha et al.); (iii) basis of flowering time variations (Fang et al.); (iv) role of high-throughput phenotyping role in enlightening complex traits (Ebersbach et al.); (v) flower color and pigment formation (Li et al.; Hao et al.; Yang et al.); (vi) harvest index process and mechanisms (Zhang et al.); (vii) factors responsible for control of seed development stages and yield in special eco-environments (Xiong et al.); (viii) N rate-Auxin-Floral Meristem crosstalk (Hao et al.); (ix) cold resistance (Wu et al.); (x) male sterile lines role for hybrid seed production (Dong et al.); (xi) a comprehensive analysis on simple sequence repeats markers (Xu et al.); (xii) genome-wide characterization of ovate family protein gene family (Liu et al.); (xiii) relation of seed

number per silique and ovule number per ovary with ovule fertilization and seeds development (Qadir et al.); (xiv) minimization of erucic acid in seed oil (Gill et al.); (xv) loci of clubroot-resistant genes (Wang et al.); and (xvi) functional characterization of the receptor-like proteins under the stimulus of *Sclerotinia sclerotiorum* (Li et al.).

Brief highlights and the major outcomes of the studies on *Brassica napus*, *B. rapa*, *B. juncea*, and *B. oleracea* included in this Research Topic are outlined below.

Brassica napus

As the third-largest oilseed crop worldwide, rapeseed (*Brassica napus*) exhibits significantly high oil production efficiency. *B. napus* has evolved from the double diploidization of *B. rapa* and *B. oleracea* through interspecific hybridization. There has been a continuous search for factors contributing to improving *B. napus* yield. To this end, silique-related traits are of great significance. Analyzing the 120 consensus quantitative trait loci (QTLs) across multiple environments, 23, 25, 29, 22, and 21 consensus QTLs were identified for five silique-related traits (seed number per silique, SPS; silique length, SL; silique breadth, SB; silique thickness, ST; silique volume, SV), respectively (Zhao et al.). Interestingly, seed number per silique (SNPS) largely rely on the ovule number per ovary (ONPO), the proportion of ovules to be fertilized and the proportion of fertilized ovules to develop into seeds. ONPO widely varies in *B. napus*; however, underlying genes and mechanisms were unveiled and found to involve 18 novel association loci and six candidate genes (*BnaA03g14600D*, *BnaA03g33420D*, *BnaA06g08920D*, *BnaA06g13210D*, *BnaC01g25840D*, and *BnaC03g16210D*) (Qadir et al.). *B. napus* suffers from low genetic diversity; hence, exploitation of diverse genetic resources is imperative to develop locally adapted, high yielding, and stress resistant cultivars. To this end, it has been suggested that high-throughput phenotyping (HTP), an automated, scalable, non-destructive, and high-throughput imaging approach, is feasible for phenotyping studies on complex traits such as drought stress resistance, flowering characteristics (e.g., timing and volume), and canopy architecture traits (e.g., raceme branch numbers) in spring-type *B. napus* lines. These were selected as founders for the development of a large, spring-type *B. napus* Nested Association Mapping (NAM) population called "SKBnNAM" (Ebersbach et al.). The yield-loss in *B. napus* in some regions of the world also involves early frost-caused early flowering (maturation), which is often associated with flowering time (FT). Three genes (*BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*) were identified and argued to be major determinants for FT-variation of two elite *B. napus* accessions (616A and R11). Moreover, exploration of heterosis in FT-genes revealed a relation between FT and plant yield with hybrid near-isogenic lines (NILs) (Fang et al.).

In Brassica species, the flower color is among the most important traits contributing to pollen transmission in nature, and ornamental and landscaping purposes. Though cloned in tomato, maize, melon, and *Arabidopsis thaliana*, report on carotenoid isomerase (CRTISO) gene, responsible for converting the yellow-colored prolycopene into the red-colored all-trans lycopene in the carotenoid synthesis pathway is lacking in *B. naups*. Notably, as the most simple and efficient sequence-specific nucleases (SSNs), CRISPR/Cas9 (Li et al.) helped edit two copies of the carotenoid isomerase gene (*BnaCRTISO*) in *B. napus* (*BnaA09.CRTISO* and *BnaC08.CRTISO*), recovering the mutation phenotype of creamy white petals and yellowish leaves, and establishing a correlation between the carotenoid pathway and flavonoid synthesis pathway, so far not well known in most plants. Information is meager on the major mechanism underlying the involvement of the anthocyanin pathway (crucial in plant color development, ranging from pink to blue and purple) in *B. napus* petal color formation. An investigation employing metabolomics and RNA-seq studies on two different stages of unopened petals of red, pure white, and yellow petal *B. napus* lines identified *BnaA03.ANS* as the gene involved in *B. napus* petal color control (Hao et al.).

Harvest index (HI) is a complex agronomic trait and an economically critical value, representing the ratio of harvested seed weight to total aboveground biomass weight. Unfortunately, *B. napus* exhibit values of HI that are much lower than in other major crops, and the molecular-genetic regulatory network underlying HI in *B. npus* is largely unknown. Interestingly, studies on the *B. napus* accessions YC24, YC52, and YC46, exhibiting differential HI, mRNA, and small RNA sequencing revealed the role of transporter activity-related genes in enhancing HI under good cultivation environments (Zhang et al.). The development of seeds (which comprises three major stages: sugar mobilization, sequential surges in amino acid, lipid, and storage protein synthesis) is largely modulated by environmental factors in a special eco-environment exhibiting differential sunshine duration, temperature, and water. Notably, when cultivated under such a special eco-environment, *B. napus* (Qingza 5, a spring rapeseed variety)-grain development was found to be ahead of schedule, which lasted for a long time, and led to a high-yield (Xiong et al.). The production of *B. napus* is significantly modulated by conversion from vegetative to reproductive growth, which in turn is determined by the acceleration of the differentiation of floral meristem (FM) from shoot apical meristems (SAM). SAM differentiation can be regulated by many factors including phytohormones (such as auxin), and environmental and agronomic practices such as low temperature and nutrients (such as N). Interestingly, a high N rate can accelerate the initiation of FM differentiation, and involve auxin, which in turn involves both auxin biosynthesis genes (*ASA1/ASA2*, *IGS*, *TSA1*, *TSB1*, *CYP79B*, *NIT1/NIT2*, and *AMI1*) and also indole acetaldehyde oxime pathway as the major

pathway for auxin biosynthesis (Hao et al.). *B. napus* has experienced whole genome duplication and shares significant homology with *Arabidopsis thaliana*, known to possess multiple copies of genes of ovate family proteins (OFPs; AtOFP1, AtOFP4, AtOFP5). However, B. napus was not been reported to exhibit any function of OFPs. The genome-wide characterization of the OFP gene family associated with the number of seeds per silique (NSPS) in B. napus helped in the identification of both the BnOFP gene family at the genomic level and a new locus *BnOFP13_2*, which was significantly correlated with NSPS (Liu et al.). Among the major multiple external stimuli, *Sclerotinia sclerotiorum* has emerged as an important Brassica pathogenic fungi also known to remarkably affect the yield and quality of *B. napus*. A comparison of the transcriptional profiles of genes of receptor-like proteins (RLPs, an indispensable constituent in the first layer of defense, and typically with tandem leucine-rich repeats) with that of the functional assigned *AtRLP* genes revealed *AtRLP10* (CLV2, AT1G65380.1) as a regulator of plant meristem and organ development, and *S. sclerotiorum*-exerted stimulus-caused decrease in transcript abundance of *BnaC02g45200D* and *BnaA02g12070D* (Li et al.).

Brassica rapa

One of the three diploid ancestors of *B. napus* and *B. juncea*, *B. rapa* is highly diverse, and of nutritional and economic importance as it is widely cultivated worldwide as an oil and vegetable crop species. Few studies have reported on the genes underlying the dark yellow flower trait in Chinese cabbage (*B. rapa* L. ssp. *Pekinensis*). To this end, *Bra037130* (*BraA09.ZEP*) (which encodes a zeaxanthin epoxidase) has been argued to be the most likely candidate gene for *Br-dyp1* involved in the epoxidation from zeaxanthin to violaxanthin in *B. rapa* L. ssp. Pekinensis (Yang et al.). The health and yield of *B. rapa* (winter rapeseed) are lagging due to chronic low temperatures in winter in some regions of the world such as northwest China. Elucidation of candidate genes associated with cold resistance in *B. rapa* has been limited by currently existing genomes. Taking into account a freeze-tolerant winter turnip variety "Longyou-7" (LY7), assembly of its high-quality genome (using PacBio HIFI reads) and construction of a graph-based pan-genome by combining LY7-genome with the other 22 *B. rapa* accessions led to the identification of *HDG1* and *BrANS3* as the two genes associated with cold tolerance (Wu et al.). Breeding of male sterile lines is crucial for the hybrid seed production and commercialization of *B. rapa* L. ssp. *Pekinensis*, which is a typical cross-pollinated Brassica crop with obvious heterosis. Interestingly, three allele male sterile mutants (*msm2-1/2/3*) were identified in an ethyl methane sulfonate (EMS)-mutagenized *B. rapa* L. ssp. *Pekinensis*, where pollen abortion was obtained due to non-synonymous base-pair mutations in BrMS1 (*BraA10g019050.3C*), and *AtMS1* orthologs gene (relevant to fertility and pollen development regulation) was cloned first time in EMS-mutagenized *B. rapa* L. ssp. *Pekinensis* (Dong et al.).

The life of most Brassica species (including *B. napus*, *B. oleracea*, and *B. rapa*) is endangered by a highly contagious soil-borne disease, clubroot (caused by *Plasmodiophora brassicae*). Exploration is imperative to identifying new resources for resistance genes owing to reports on the loss of resistance over time and the presence of various pathotypes for clubroot-resistant (CR) pathogens from different regions. Notably, DingWen (DW) identified a unique *B. rapa* material that shows different resistance capabilities from H5R (PbBa8.1) and Huayouza62R (CRb) in *B. napus*. Moreover, the peculiar resistance in DW most likely involves the CR candidate genes *BraA08g039211E* and *BraA08g039212E* (encoding TIR-NBS-LRR proteins), and *BraA08g039193E* (encoding an RLP protein), which could also act as a new gene resource for CR breeding in the future (Wang et al.).

Brasica juncea

Indian mustard (*B. juncea*) is a major oilseed crop in the south Asia region, including the Indian sub-continent. Unfortunately, *B. juncea* seed oil exhibits a very high concentration (≈50%) of erucic acid, which, if continuously consumed, may cause cardiac lipidosis. KASPar assay, when developed to select low erucic acid and to track recessive alleles in their heterozygous state, is single step, has a high throughput, and is robust, codominant, and more cost-effective than gel-based markers (Gill et al.). The heterotic quantitative trait loci (QTL) has been mapped and a large number of epistatic interactions were identified for seed yield and three yield-component traits in *B. juncea* using doubled haploids (DH) and the corresponding backcross lines with their midparent heterosis data. However, the molecular mechanism underlying the complex phenomenon, particularly in *B. juncea* is not well understood. Fortunately, the complex genetic basis of heterosis of yield-influencing traits (plant architecture, flowering, and silique- and seed-related traits) was unveiled in *B. juncea* and argued to involve a large number of additive QTLs (695) and loci (637 epistatic loci). It also exhibited cumulative effects of dominance, overdominance, and a large number of epistatic interactions (Aakanksha et al.).

Brassica oleracea

As one of the most critical cruciferous vegetables, cabbage (*B. oleracea* L. var. *capitata*) has been widely cultivated all over the world since it contains favorable components for human

health. Notably, compared with the development of simple sequence repeats (SSR) markers in other crops, SSR markers, those developed from *B. oleracea* L. var. *capitata* reference genome are lacking. Additionally, despite the contribution of RNA-Seq data to in silico generation, a comprehensive analysis of SSRs within *B. oleracea* L. var. *capitata* genome is not available. To fill the highlighted knowledge-gaps, 64,546 perfect and 93,724 imperfect SSR motifs were identified in the 0.5Gb of *B. oleracea* L. var. *capitata* genomic sequence (TO1000), which was mined using a whole-genome bioinformatics survey (Xu et al.).

Conclusions and outlook

The contributions to volume II of the Research Topic '*The Brassicaceae - Agri-Horticultural and Environmental Perspectives*' discussed the molecular-genetic insights into agri-horticultural aspects that can help in devising breeding approaches in various species of the genus Brassica for yield, flowering time variations, flower color, and pigment formation. The collection discusses the factors involved in seed development stages and yield in special eco-environments, N rate-Auxin-Floral Meristem crosstalk, resistance to cold and pathogens, hybrid seed production, and minimization of erucic acid in seed oil. However, no contributions discuss the molecular-genetic aspects of abiotic stress tolerance and the environmental perspectives of the Brassicaceae family members. Additionally, there is a large number of existing research on the contribution of the members of the plant family Brassicaceae in the control of varied environmental issues. These highlighted aspects will be further explored in future volume(s) of this Research Topic.

Author contributions

NA prepared the first draft of the manuscript. SG OD, JJ, and NT read and revised the manuscript, and all authors approved the final version for publication.

Funding

NA gratefully acknowledges the financial support from the Department of Biotechnology (DBT), Government of India for financial assistance in the form of the DBT-Ramalingaswami Re-Entry Fellowship (BT/HRD/35//02/2006; BT/RLF/Re-Entry/18/2015). SG and NT also acknowledge partial financial support from the University Grants Commission (UGC) and Council of Scientific and Industrial Research (CSIR), Government of India, New Delhi.

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fpls.2021.766271

QTL Analysis of Five Silique-Related Traits in *Brassica napus* L. Across Multiple Environments

*Xiaozhen Zhao[1,2†], Kunjiang Yu[3†], Chengke Pang[1,2], Xu Wu[1], Rui Shi[1,2], Chengming Sun[1], Wei Zhang[1], Feng Chen[1], Jiefu Zhang[1,2] and Xiaodong Wang[1]**

[1] Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Key Laboratory of Cotton and Rapeseed, Ministry of Agriculture and Rural Affairs, Nanjing, China, [2] State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, [3] College of Agriculture, Guizhou University, Guiyang, China

As an important physiological and reproductive organ, the silique is a determining factor of seed yield and a breeding target trait in rapeseed (*Brassica napus* L.). Genetic studies of silique-related traits are helpful for rapeseed marker-assisted high-yield breeding. In this study, a recombinant inbred population containing 189 lines was used to perform a quantitative trait loci (QTLs) analysis for five silique-related traits in seven different environments. As a result, 120 consensus QTLs related to five silique-related traits were identified, including 23 for silique length, 25 for silique breadth, 29 for silique thickness, 22 for seed number per silique and 21 for silique volume, which covered all the chromosomes, except C5. Among them, 13 consensus QTLs, one, five, two, four and one for silique length, silique breadth, silique thickness, seed number per silique and silique volume, respectively, were repeatedly detected in multiple environments and explained 4.38–13.0% of the phenotypic variation. On the basis of the functional annotations of *Arabidopsis* homologous genes and previously reported silique-related genes, 12 potential candidate genes underlying these 13 QTLs were screened and found to be stable in multiple environments by analyzing the re-sequencing results of the two parental lines. These findings provide new insights into the gene networks affecting silique-related traits at the QTL level in rapeseed.

Keywords: *Brassica napus*, recombinant inbred line, silique-related traits, QTL mapping, candidate genes

OPEN ACCESS

Edited by:
Kun Lu,
Southwest University, China

Reviewed by:
Lunwen Qian,
Hunan Agricultural University, China
Wei Qian,
Chinese Academy of Agricultural Sciences (CAAS), China

****Correspondence:***
Xiaodong Wang
xdwang120@163.com

† These authors have contributed equally to this work

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

Received: *28 August 2021*
Accepted: *06 October 2021*
Published: *23 November 2021*

Citation:
Zhao X, Yu K, Pang C, Wu X, Shi R, Sun C, Zhang W, Chen F, Zhang J and Wang X (2021) QTL Analysis of Five Silique-Related Traits in Brassica napus L. Across Multiple Environments. Front. Plant Sci. 12:766271. doi: 10.3389/fpls.2021.766271

INTRODUCTION

Rapeseed (*Brassica napus*, AACC, 2n = 38), a major oil crops worldwide, evolved from the double diploidization of *Brassica rapa* (AA, 2n = 20) and *Brassica oleracea* (CC, 2n = 18) through interspecific hybridization (Murphy, 1999). *B. napus* is an oil crop with the highest oil production efficiency and is currently an important source of edible vegetable oil in China (Wang, 2010). Because of the high protein content of rapeseed meal, rapeseed is also a high-quality source of animal feed. Additionally, rape straw undergoes an efficient biomass digestion and can be used as a raw material for bio-energy production (Wang X. et al., 2016).

The global population, and thus, food demand, continues to grow; however, the amount of cultivated land is decreasing owing to a variety of factors, such as human activities and climate change. Therefore, increasing the production of rapeseed is currently an urgent requirement and

major goal of global rapeseed producers. Yield is an extremely complex trait, and the product of a series of developmental and physiological processes (Van Camp, 2005). Rapeseed yield is determined by three constituent factors: seed weight, silique number and seed number per silique (SPS) (Clarke and Simpson, 1978). The silique, as an important photosynthetic organ of rapeseed, is closely related to the final grain yield (Allen et al., 1971). Silique-related traits, such as silique length (SL), silique breadth (SB), silique thickness (ST) and silique volume (SV), affect the appearance and morphology of silique, and ultimately affect the production and yield of photosynthetic substances (Ferrándiz et al., 1999). Thus, siliques and their related traits are considered the major contributing factors for increasing rapeseed yield. Understanding their genetic bases is of great significance for breeding high-yield rapeseed.

The silique-related traits of rapeseed are all complex traits controlled by quantitative trait loci (QTLs), which are also easily affected by the environment. QTL mapping has been successfully applied to the genetic analyses of the quantitative traits of a variety of crops. In rapeseed, a large number of QTLs, such as those affecting plant height (Wang Y. et al., 2016; Dong et al., 2021), branch angle (Wang H. et al., 2016; Wang et al., 2019) and flowering time (Yu et al., 2019; Xu et al., 2021), have been genetically mapped. In siliques, SL and SPS are the most important traits. At present, more than 100 QTLs for SL, along with more than 200 QTLs for SPS, have been detected in different mapping populations of *B. napus*, and QTLs controlling SL and SPS are distributed on all 19 chromosomes, explaining 1.9–65.6% and 0.78–57.8% of the phenotypic variance (PV), respectively (Chen et al., 2007, 2011; Radoev et al., 2008; Shi et al., 2009; Zhang et al., 2011; Yang et al., 2012, 2016, 2017; Cai et al., 2014, 2016; Qi et al., 2014; Fu et al., 2015; Wang X. et al., 2016; Luo et al., 2017; Zhu et al., 2020). Recently, Li et al. (2015) successfully isolated *qSS.C9*, the main locus controlling SPS in *B. napus* that had been discovered by Zhang et al. (2012). *qSS.C9* encodes the predicted protein BnaC9.SMG7b. In *Arabidopsis*, SMG7 controls female fertility and then SPS by regulating the progression from anaphase to telophase in the second meiotic division (Riehs et al., 2008). As in *Arabidopsis,* BnaC9.SMG7b determines the formation of functional female gametophytes (FGs) by influencing the progression through meiotic anaphase II. Plants with *BnaC9.SMG7b* deletions exhibit reduced SPS and SL values, which is caused by the reduced ovule number and shorter siliques because of defects in the development of functional FGs. In addition, two major QTL for SL have been successfully cloned on the A9 chromosome of *B. napus* (Liu et al., 2015; Shi et al., 2019). The target genes of these two QTLs were identified as *BnaA9.ARF18* and *BnaA9.CYP78A9* through a fine-mapping analysis. *BnaA9.ARF18* is a homolog of the *Arabidopsis* auxin response factor 18 (*ARF18*) that is expressed differentially in various tissues, including root, leaf, stem, bud and ovule, but mainly in the silique wall. A transcription analysis has shown that *ARF18* regulates cell elongation in the silique wall and then SL by acting through the auxin-response pathway (Liu et al., 2015). *BnaA9.CYP78A9* is an ortholog of *Arabidopsis CYP78A9,* which regulates reproductive development and floral organ size (Sotelo-Silveira et al., 2013). *BnaA9.CYP78A9* affects SL in rapeseed by promoting cell elongation in silique valves during silique growth and development (Shi et al., 2019). Although a few silique genes, such as *BnaC9.SMG7b*, *BnaA9.ARF18* and *BnaA9.CYP78A9*, have been analyzed in *B. napus*, the genetic mechanisms behind silique-related traits, especially those other than SPS and SL, are far from understood.

The formation of the SPS is closely related to other silique characteristics (SL, SB, ST, and SV), and SPS is significantly positively correlated with SL (Zhang et al., 2011; Wang X. et al., 2016; Yang et al., 2017). As an important silique-related morphological trait, SL also plays a vital role in seed yield, and it has long been used as an indirect selective indicator in breeding for improved seed yields in *B. napus*. To date, there are extremely limited QTL mapping studies of SB, ST, and SV in *B. napus* (Wang X. et al., 2016), and more research is needed to determine the relationships among them and their effects on the yield, which is very important for increasing the efficiency levels of breeding programs.

In the present study, a recombinant inbred line (RIL) population containing 189 lines was used to investigate the QTLs for five silique-related traits in seven environments. The objectives were to identify (i) QTLs associated with SL, SPS, SB, ST, and SV across multiple environments and (ii) candidate genes underlying the QTLs that could be stably detected based on the re-sequencing information of two parental lines. This study will lay a good foundation for studying the molecular mechanisms of silique-related traits and increasing yield through rapeseed breeding.

MATERIALS AND METHODS

Plant Materials

An RIL population containing 189 lines was constructed by a cross between "APL01" (female parent) and "Holly" (male parent), and it was named the AH population (Wang et al., 2015). The two parental lines showed diversity in silique-related traits. The AH population was previously used for high-density SNP genetic map construction (Wang et al., 2015). The map contains 2,755 SNP-bins, including 11,458 SNP markers and 57 simple sequence repeats, spanning a genetic distance of 2,027.53 cM, with an average distance of 0.72 cM between markers. This population was previously used for QTL mapping and the analysis of apetalous characteristics (Wang et al., 2015), seed fatty acid composition (Chen et al., 2018) and seed-related traits (Sun et al., 2018). In this study, it was used for the QTL mapping of five silique-related traits.

Field Experiment and Trait Measurement

The AH population, and the parental lines, were grown in four locations in China in 2014–2016. Yangling, Shaanxi Province was planted in September 2015, harvested in May 2016 and recorded as 15YL; Sunan, Gansu Province, was planted in April and harvested in September 2016 and recorded as 16GS; Dali, Shaanxi Province was planted in September 2014 and 2016, harvested in May 2015 and 2017, and recorded as 14DL and 16DL, respectively; and Nanjing, Jiangsu Province was

planted in September 2014, 2015 and 2016, harvested in the May of the following years and recorded as 14NJ, 15NJ, and 16NJ, respectively. The field planting of the parents and AH population followed a randomized complete block design with two replications. Each repetition contained two rows with an average spacing of 40 cm between rows and 20 cm between individual plants. The planting, management and harvesting of field materials followed local field-breeding practices.

At the mature stage, three plants growing uniformly from each line were selected, and three well-developed siliques of the first branch adjacent to the main inflorescence from each plant were collected to analyze the silique-related traits (Wang X. et al., 2016). Here, five silique-related traits were investigated: (1) SL, the average length of nine sampled siliques; (2) SB, the widest horizontal length of the space occupied by the silique; (3) ST, the thickness at the same position as SB; (4) SV, the volume measured using the drainage method, which takes the difference between the volume reading of the water after silique immersion and the volume reading of the water prior to the immersion; and (5) SPS, the average seed number of nine sampled siliques. A Vernier caliper was used to measure SL, SB and ST, and a volumetric cylinder was used to measure SV.

Statistical Analysis

SPSS 22.0 software (SPSS Inc., Chicago, IL, United SA) was used for descriptive statistical and correlation analyses of each trait in each environment, and the mean value of the seven environments for each trait was used to carry out correlation analyses. The coefficient of variation was calculated as σ/μ, where σ represents the standard deviation and μ represents the average. The broad-sense heritability (h^2) was obtained using the R package lme4 (Merk et al., 2012). The computational formula for broad-sense heritability was as follows: $h^2 = \sigma^2_G/(\sigma^2_G + \sigma^2_{GE}/e + \sigma^2_e/re)$, where σ^2_G, σ^2_{GE}, and σ^2_e, respectively, represent genotypic variance, the interaction variance of genotype-by-environment and error variance, n represents the number of environments and r represents the number of replications (Sun et al., 2016).

QTL Mapping and Integration

The composite interval mapping procedure of the Windows QTL Cartographer 2.5 was used to detect the QTLs associated with the five silique-related traits (Wang et al., 2012). The window size was set to 10 cM, a walking speed of 2 cM was selected, and five markers were set as background cofactors. The permutation analysis with 1,000 repetitions ($P = 0.05$) was performed, and a LOD threshold (2.5–2.8) was used to determine the existence of a QTL and named as "identified QTL." The QTL intervals were established by 2-LOD as approximately 95% QTL confidence intervals (CIs), which is automatically generated by Windows QTL Cartographer 2.5. Identified QTLs for each trait have undergone two rounds of meta-analysis using BioMercator 2.1 software, which can be used to decide the best fitting QTL in accordance with the Akaike criterion (Arcade et al., 2004). The QTL meta-analysis was performed following the description of Wang X. et al. (2016). In the first round, identified QTLs for specific traits with overlapping CIs detected in two or more environments were integrated and named as "consensus QTLs." In the second round, consensus QTLs for different traits located in the same chromosomal region were integrated and named as "unique QTLs" (Chen et al., 2018). The designation of the QTL refers to the method of McCouch et al. (1997) with modifications. The name of identified QTL starts with its abbreviation "*iq*," followed by the environment/trait abbreviation and linkage group (A1–10 and C1–9). When there are multiple QTLs within the same linkage group, the identified QTLs are numbered in accordance with their physical locations. Similarly, the names of consensus QTLs and unique QTLs start with their abbreviations "*cq*" and "*uq*," respectively.

Prediction of Underlying Candidate Genes for SL, SB, ST, SPS, and SV

The Illumina HiSeq 2500 platform (Illumina, Inc., San Diego, CA, United States) was used to re-sequence the parental APL01 and Holly genomes with a sequencing depth of 30 × coverage in our previous study (Yu et al., 2021). A total of 74.89 G bp clean data were obtained, with Q30 reaching an average of 92.0%. For candidate gene prediction the following steps were taken: First, the SNP/Indel loci with differences and homozygous between the parents were detected. The clean reads of APL01 and Holly were aligned against *B. napus* reference genome "Darmor-*bzh*" (Chalhoub et al., 2014) using BWA software (Li and Durbin, 2009). The SAMtools command "mpileup" was used to genotypes calling (Li et al., 2009). The locus were screened and retained with criteria: (1) Depth $\geq$ 6; (2) Quality $\geq$ 20; (3) Allele mutation ratio of homozygous loci $<$0.1 or $>$0.9. After filtration, the genotypes of the two parents were combined by the Bcftools command "merge" (Danecek and McCarthy, 2017), and the SNP/Indel loci with differences and homozygous between the parents were retained; Second, the probe sequences of SNPs on the genetic map were used to map markers on both sides of the QTL interval to the physical position of the reference genome "Darmor-*bzh*" using BLAST (E value $\leq$ 1e-10) (Chalhoub et al., 2014), and the variation loci within the QTL interval were found using the results of the first step; and Third, on the basis of the functional annotations of *Arabidopsis* homologous genes and previously reported silique-related genes, candidate genes for silique-related traits of rapeseed were screened from the mutant loci of these QTLs.

RESULTS

Phenotypic Analysis of Five Silique-Related Traits

The phenotypic performance and h^2 estimates for the five silique-related traits in the two parental lines and AH population are presented in **Table 1**. The five silique-related traits of the two parents, "APL01" and "Holly," differed significantly in most of the investigated environments, and only two environments, 16DL and 16NJ, showed no significant differences in ST. Compared with "Holly," "APL01" had significantly greater SB, ST, SV,

TABLE 1 | Statistical analysis of five silique-related traits for two parents and AH populations in 7 environments.

		Parental line		Pt-test	RILs					
Trait	**Treatment**	**APL01 (mean ± SD)**	**Holly (mean ± SD)**		**Mean ± SD**	**Range**	**CV**	**Skewness**	**Kurtosis**	h^2 **(%)**
SL (cm)	14DL	6.02 ± 0.58	7.14 ± 0.48	6.77E-04**	6.23 ± 0.76	4.0-9.0	0.12	0.08	−0.11	81.10
	14NJ	6.33 ± 0.27	6.69 ± 0.22	1.06E-02*	6.20 ± 0.75	3.5-8.6	0.12	0.02	0.58	
	15NJ	5.92 ± 0.38	6.47 ± 0.36	9.69E-03*	6.34 ± 0.76	3.9-9.3	0.12	0.05	−0.04	
	15YL	5.30 ± 0.65	5.86 ± 0.31	4.43E-02*	6.15 ± 0.78	4.0-10.7	0.13	0.72	2.72	
	16DL	5.41 ± 0.42	6.04 ± 0.32	4.00E-03**	6.21 ± 0.81	3.7-9.1	0.13	−0.02	0.20	
	16GS	6.12 ± 0.56	7.01 ± 0.52	5.00E-03**	6.38 ± 0.75	4.2-9.0	0.12	0.18	0.02	
	16NJ	5.60 ± 0.19	6.62 ± 0.55	1.41E-04**	6.19 ± 0.79	1.5-10.8	0.13	0.37	2.42	
SB (mm)	14DL	5.56 ± 0.08	4.34 ± 0.21	1.58E-03**	4.58 ± 0.43	3.2-6.6	0.09	0.39	1.23	82.45
	14NJ	5.55 ± 0.13	4.51 ± 0.40	2.46E-02*	4.84 ± 0.47	3.2-6.0	0.10	−0.05	0.40	
	15NJ	5.63 ± 0.20	4.44 ± 0.24	5.65E-03**	5.01 ± 0.50	3.5-7.0	0.10	0.20	−0.01	
	15YL	5.57 ± 0.13	4.42 ± 0.23	3.44E-03**	4.93 ± 0.46	3.7-6.6	0.09	0.36	0.24	
	16DL	5.36 ± 0.31	4.15 ± 0.21	9.96E-03**	4.69 ± 0.45	3.6-6.0	0.10	0.12	−0.15	
	16GS	5.56 ± 0.19	4.49 ± 0.21	5.77E-03**	4.93 ± 0.47	3.6-7.2	0.10	0.74	2.09	
	16NJ	5.08 ± 0.17	4.04 ± 0.45	3.00E-03**	4.91 ± 0.57	3.3-7.1	0.12	0.39	0.36	
ST (mm)	14DL	3.81 ± 0.13	3.36 ± 0.11	2.19E-02*	3.34 ± 0.28	2.6-4.2	0.08	0.29	0.25	76.66
	14NJ	3.75 ± 0.18	3.33 ± 0.02	2.68E-02*	3.42 ± 0.33	2.5-4.5	0.10	0.28	0.15	
	15NJ	3.84 ± 0.06	3.50 ± 0.10	1.38E-02*	3.51 ± 0.33	2.7-5.3	0.09	0.46	0.84	
	15YL	3.67 ± 0.03	3.46 ± 0.09	3.14E-02*	3.48 ± 0.30	2.4-4.7	0.09	0.22	0.52	
	16DL	3.70 ± 0.09	3.37 ± 0.14	5.43E-02	3.48 ± 0.29	2.7-4.6	0.08	0.40	0.73	
	16GS	3.70 ± 0.07	3.49 ± 0.06	3.24E-02*	3.52 ± 0.31	2.6-4.8	0.09	0.17	0.52	
	16NJ	3.72 ± 0.11	3.57 ± 0.14	3.27E-01	3.41 ± 0.42	2.2-5.8	0.12	0.86	1.99	
SV (ml)	14DL	0.88 ± 0.10	0.56 ± 0.01	1.29E-02*	0.68 ± 0.14	0.3-1.2	0.21	0.46	0.52	73.10
	14NJ	0.92 ± 0.08	0.61 ± 0.07	7.30E-05**	0.70 ± 0.17	0.3-1.2	0.24	0.34	−0.06	
	15NJ	0.98 ± 0.03	0.66 ± 0.03	6.08E-04**	0.80 ± 0.15	0.4-1.5	0.19	0.80	1.84	
	15YL	0.98 ± 0.06	0.60 ± 0.03	9.32E-04**	0.79 ± 0.16	0.5-1.5	0.20	1.18	3.67	
	16DL	0.91 ± 0.03	0.41 ± 0.07	6.78E-04**	0.66 ± 0.14	0.3-1.3	0.20	0.60	0.86	
	16GS	0.99 ± 0.08	0.60 ± 0.00	2.41E-03**	0.86 ± 0.14	0.5-1.7	0.17	1.02	3.77	
	16NJ	0.91 ± 0.07	0.50 ± 0.10	5.80E-5**	0.64 ± 0.18	0.2-1.3	0.29	0.49	0.45	
SPS	14DL	25.0 ± 2.8	19.7 ± 2.1	4.91E-04**	22.5 ± 4.23	2-37	0.19	−0.17	0.38	73.61
	14NJ	27.9 ± 2.8	22.4 ± 2.1	4.64E-04**	21.9 ± 4.25	2-36	0.19	−0.27	0.69	
	15NJ	28.0 ± 4.8	23.6 ± 1.8	2.65E-02*	23.1 ± 5.00	5-38	0.22	−0.25	0.00	
	15YL	22.7 ± 3.0	18.3 ± 2.7	7.75E-03**	22.4 ± 4.57	9-42	0.20	0.04	0.00	
	16DL	21.4 ± 1.9	17.9 ± 1.5	9.02E-04**	21.9 ± 4.67	2-38	0.21	−0.01	0.25	
	16GS	30.6 ± 3.5	26.8 ± 2.3	2.20E-02*	24.4 ± 4.05	2-40	0.17	−0.13	0.66	
	16NJ	25.9 ± 2.64	15.7 ± 2.21	3.0E-7**	22.3 ± 4.28	2-35	0.19	−0.09	−0.13	

SL, SB, ST, SPS and SV indicate the traits silique length, silique breadth, silique thickness, seed number per silique and silique volume, respectively.
DL, Dali; NJ, Nanjing; YL, Yangling; GS, Gansu; 14, 15, and 16 indicate the years 2014, 2015 and 2016, respectively.
***Significant at the 0.01 probability level.*
**Significant at the 0.05 probability level.*
CV, Coefficient of variation.
h^2, Broad-sense heritability.

and SPS values, whereas "Holly" had a significantly greater average SL than "APL01." The h^2 values for SL, SB, ST, SV, and SPS were 81.10, 82.45, 76.66, 73.10, and 73.61% in the AH population, respectively (**Table 1**), suggesting that the QTLs controlling these traits have large effects on breeding rapeseed to increase yield.

A wide range of variation, as well as transgressive segregation, were observed for the five silique-related traits, suggesting that alleles with positive effects were distributed in both parents (**Table 1** and **Figure 1**). In addition, the distributions of the five traits were continuous, and most skewness and kurtosis values for the distributions of these traits were <1.0 (**Table 1** and **Figure 1**), which are characteristic of the normal distribution model, indicating that the AH population is suitable for QTL mapping.

Correlation Analysis Among Five Silique-Related Traits

The correlation coefficients among the five silique-related traits were calculated (**Table 2**), and most were extremely significantly correlated with each other. For instance, highly

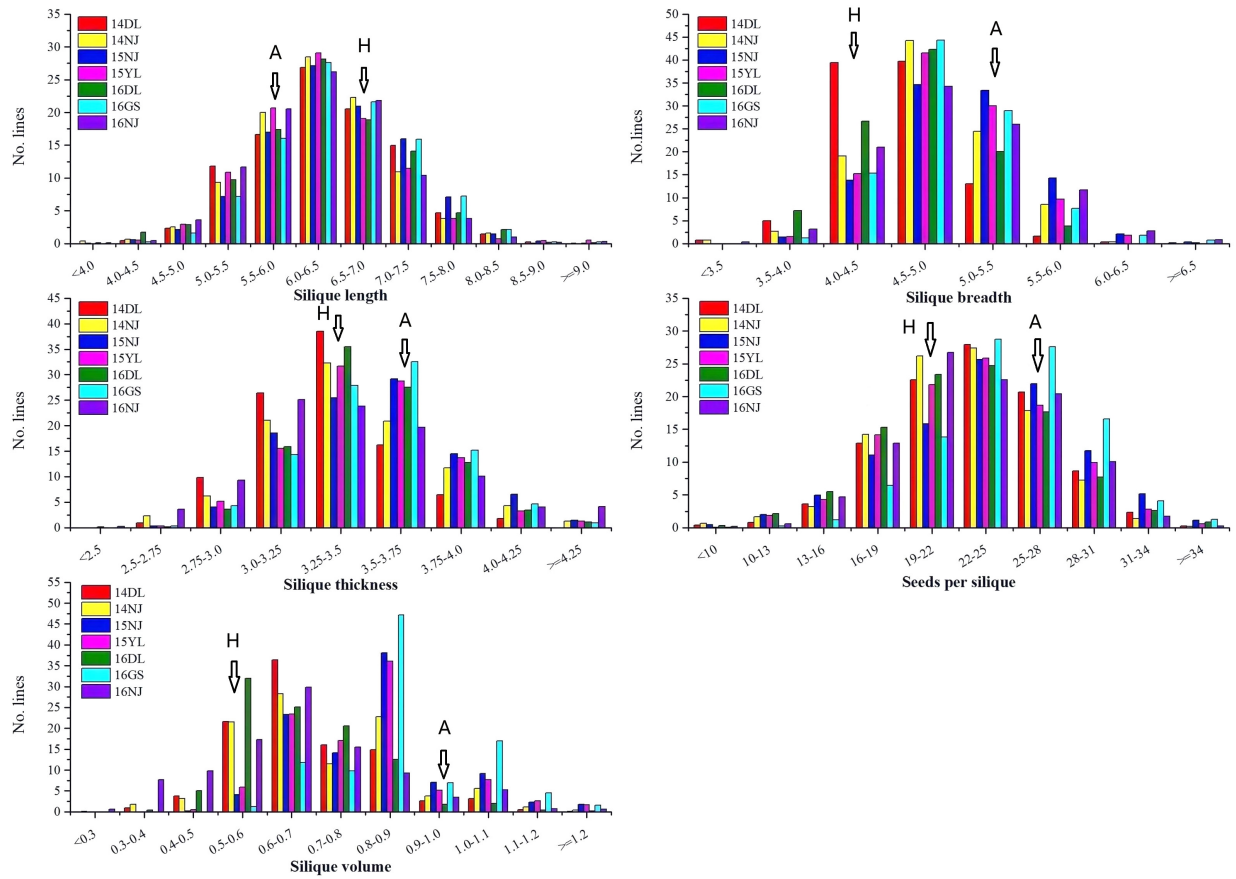


FIGURE 1 | Phenotype frequency distributions of five silique-related traits in the AH population. 14DL, 14NJ, 15NJ, 15YL, 16DL, 16GS, and 16NJ are the codes for the following different years and environments: 2014 Dali, 2014 Nanjing, 2015 Nanjing, 2015 Yangling, 2016 Dali, 2016 Gansu, and 2016 Nanjing, respectively. A represents the female parent "APL01" and H represents the male parent "Holly" of the AH population.

significant positive correlations were observed between SV and the other four silique-related traits, SB was significantly positively correlated with ST, SPS and SV, and SPS showed significant positive correlations with the other silique-related traits, except for ST. Studying the associations among SL, SB, ST, SPS, and SV is of great significance to understand silique morphogenesis and increase rapeseed yield.

QTL Mapping of Five Silique-Related Traits

Genome-wide QTL analyses were performed separately using the phenotypic data of SL, SB, ST, SPS, and SV grown in seven environments, and a total of 137 identified QTLs distributed across 18 chromosomes were obtained, which explained 4.23–15.31% of the PV (**Supplementary Additional File 1**). Among the QTLs, 24 were identified for SL, 33 for SB, 31 for ST, 27 for SPS, and 22 for SV (**Supplementary Additional File 1**). For each trait, QTLs with overlapping CIs detected in different environments were integrated into one QTL through a meta-analysis. As a result, 137 identified QTLs were integrated into 120 consensus QTLs (**Figure 2**, **Supplementary Additional Files 1**, **2**). Among them, 107, 10, 2 and 1 consensus QTL were detected in 1, 2, 3, and 4 environments, respectively. The consensus QTLs repeatedly detected in multiple environments are listed in **Table 3**.

For the 24 identified QTLs of SL, which were located on chromosomes A3, A5, A6, A7, A9, A10, C2, C3, C6, and C9, each individual QTL explained 4.8–9.2% of the PV. Two identified QTLs, *iq14DL.A10-2* and *iq15YL.A10*, with overlapping CIs were integrated into the consensus QTL *cqSL.A10-3* (**Table 3**). Among the 23 consensus QTLs after integration, all the QTLs showed positive additive effects, except *cqSL.A5-3*, suggesting that the female parent "APL01" contributed favorable alleles.

For the 33 identified QTLs of SB, each QTL accounted for 4.23–15.3% of the PV, and these QTLs were mapped onto 12 chromosomes. Finally, the 33 identified QTLs were integrated into 25 consensus QTLs, including 3, 1 and 1 QTLs stably expressed in 2, 3 and 4 environments, respectively (**Table 3**). The stable QTL *cqSB.C6-1* showed a larger effect in the DL than in the NJ environment, with PVs of 12.96 and 9.53% in 14DL and 16DL, respectively but PVs of 4.38 and 5.18% in 14NJ and 15NJ, respectively. The QTL *cqSB.C6-1* was repeatedly

detected in four environments and may be used in marker-assisted selection. Four of the five consensus QTLs (*cqSB.C1*, *cqSB.C7-2*, *cqSB.C7-4,* and *cqSB.C6-1*) repeatedly detected in different environments had positive additive effects, indicating that the alleles responsible for increasing SB were inherited from female parent "APL01."

For the 31 identified QTLs of ST, each QTL explained 4.5–9.5% of the PV and were located on ten chromosomes. The 31 QTLs were integrated into 29 consensus QTLs, including two QTLs that were repeatedly detected in two environments. QTL *cqST.A5-2* was integrated from *iq16DL.A5* and *iq16GS.A5*, whereas *cqST.A5-3* was integrated from *iq14DL.A5-1* and *iq14NJ.A5-2* (**Table 3**). These two consensus QTLs had negative additive effects, indicating that the favorable alleles were derived from the male parent "Holly."

For SPS, 27 identified QTLs were obtained, explaining 4.5–13.8% of the PV. These QTLs were located on nine chromosomes. As a result of meta-QTL analyses, 22 consensus QTLs were obtained, including 18, 3 (*cqSPS.A7-3*, *cqSPS.C3-2,* and *cqSPS.C6-2*) and 1 (*cqSPS.C3-3*) QTL that were expressed in 1, 2, and 3 environments, respectively (**Table 3**). Among the 22 consensus QTLs, only four had negative additive effects and exhibited minor effects, whereas the other 18 QTLs had positive additive effects, indicating that the female parent "APL01" contributed favorable alleles.

For SV, 22 identified QTLs were detected, with each QTL explaining 4.62–11.9% of the PV. These QTLs were distributed across 13 chromosomes. There are two identified QTLs, *iq14DL.A10-1* and *iq15NJ.A10,* with overlapping CIs that were integrated into one consensus QTL, *cqSV.A10-1* (**Table 3**), and the other QTLs were only specifically expressed in a single environment (**Supplementary Additional File 1**). Of the 21 consensus QTLs, 15 had positive additive effects, including the two QTLs, *cqSV.C7-2* and *cqSV.C7-3*, which explained 10.3 and 11.9%, respectively, of the PV in 16GS, indicating that the female parent "APL01" contributed favorable alleles.

Pleiotropic Unique QTLs for the Five Silique-Related Traits

As shown in **Table 2**, most of the five silique-related traits showed highly significant correlations. The genetic correlations may be caused by pleiotropism, in which, a single QTL affects the phenotypic variation of multiple traits. Consensus QTLs for the five silique-related traits with overlapping CIs were integrated into unique QTLs using a meta-analysis. In total, 120 consensus QTLs were integrated into 89 unique QTLs (**Supplementary Additional File 3**), including 25 QTLs that had pleiotropic effects on two to three traits (**Table 4**). These pleiotropic QTLs can partly explain the significant correlations between silique-related traits. For instance, SB had a significant positive correlation with SV, with the highest positive correlation coefficient (0.718) among the five silique-related traits (**Table 2**). Five unique QTLs simultaneously controlled SB and SV, in which *uqC6-4*, *uqC7-7*, *uqC7-8*, and *uqC8* had positive additive effects, while *uqA10-8* had a negative effect on both SB and SV (**Table 4**).

TABLE 2 | Phenotypic correlations among five silique-related traits in the AH population.

Trait	SL	SB	ST	SPS	SV
SL	1				
SB	0.040	1			
ST	−0.026	0.489**	1		
SPS	0.571**	0.268**	−0.027	1	
SV	0.554**	0.718**	0.582**	0.498**	1

*SL, SB, ST, SPS, and SV indicate the traits silique length, silique breadth, silique thickness, seed number per silique, and silique volume, respectively. **Significant at the 0.01 probability level.*

Prediction of Candidate Genes for SL, SB, ST, SPS, and SV

To identify candidate genes related to SL, SB, ST, SPS, and SV, 13 consensus QTLs for the five traits that were stably expressed in multiple environments (**Table 3**) were analyzed. As a result, 2–271 genes underlying the CIs of the 13 QTLs were found (**Supplementary Additional File 4**). We analyzed these 2–271 genes one by one based on the functional annotations of *Arabidopsis* homologous genes and previously reported silique-related genes. Twelve genes underlying 12 consensus except for *cqSV.A10-1* were found to have known functions related to the silique-related traits.

The CI of *cqSL.A10-3* contained the putative gene *SMG7* (*BnaA10g15730D*), which have been successively cloned by Li et al. (2015). The CI of *cqSB.A10-2* contained the putative genes *LNG1* (*BnaA10g18650D*) and *CNGC18* (*BnaA10g19030D*), *lng1* dominant mutant displayed longitudinally elongated and transversely narrowed in cells of the siliques (Lee et al., 2006), the *CNGC18* point mutations resulted in shorter siliques, reduced male fertility and fewer seeds per silique (Gao Q. et al., 2016). The CI of *cqSB.C1* contained the putative gene *STY2* (*BnaC01g02360D*), the *sty2* mutation prevents siliques from elongating and showed short silique in *Arabidopsis* (Kuusk et al., 2002). The CI of *cqSB.C7-2* contained the putative gene *KINβγ* (*BnaC07g33610D*), its mutant displayed defects in organogenesis and growth, including shorter stature and silique (Gao X. et al., 2016). The CI of *cqSB.C7-4* contained the putative gene *ABCC13* (*BnaC07g35090D*), *ABCC13* loss function in *Arabidopsis* leads to decreased silique length and seed yield (El Guizani et al., 2014). The CI of *cqST.A5-2* contained the putative gene *ARID1* (*BnaA05g05310D*), mutation in *ARID1* showed reduced seed set and short siliques (Zheng et al., 2014). The CI of *cqST.A5-3* contained the putative gene *LAC4* (*BnaA05g06610D*), the overexpression of *miR397b* in *Arabidopsis* showed increased silique length and seed yield via modulating gene *LAC4* (Wang et al., 2014). The CI of *cqSPS.A7-3* contained the putative gene *HTH* (*BnaA07g22900D*), *hth* mutant showed reduced seed set and short siliques (Krolikowski et al., 2003). The CI of *cqSPS.C3-2* contained the putative gene *AOG1* (*BnaC03g12770D*), the siliques of *aog1* mutant were much smaller than those of wild type and had a dramatically reduced seed set (Cui et al., 2015). The CI of *cqSPS.C3-3* contained the putative gene *PYL8* (*BnaC03g15210D*), the *pyl8* T-DNA mutant

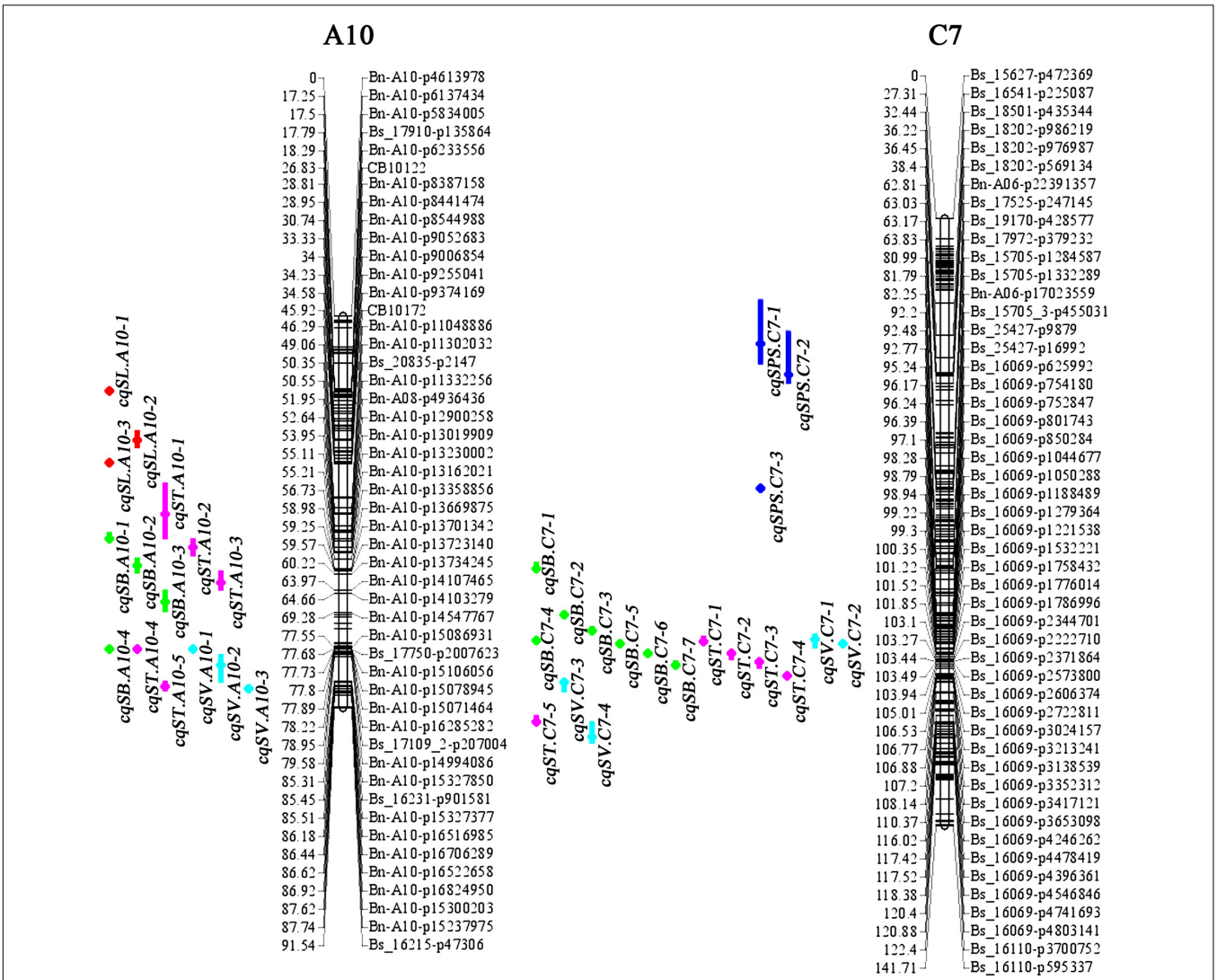


FIGURE 2 | Consensus QTL locations of silique-related traits on chromosomes A10 and C7 detected in seven different environments. Only SNPs in the confidence interval of each QTL, and the terminal two SNPs of each linkage group, are marked. The numbers listed on the left of the linkage group are distances in cM, and the locus names are on the right. Consensus QTLs for different traits are represented by different colored vertical bars: red, SL; green, SB; pink, ST; blue, SPS; turquoise, SV).

had an extremely low seed yield due to fewer siliques with lower length (Gonzalez-Guzman et al., 2012). In this study, no candidate genes were discovered in the CI of *cqSV.A10-1*. The CI of *cqSB.C6-1* and *cqSPS.C6-2* contained the putative gene *HMG1* (*BnaC06g26470D*), T-DNA insertion *hmg1* mutant showed shorter siliques and fewer seeds per silique in *Arabidopsis* (Suzuki et al., 2004).

DISCUSSION

Most agricultural production-related traits are quantitative traits controlled by multiple genes that are also sensitive to the environmental conditions (Abe et al., 2012). Therefore, QTL detection across different environments is necessary for the genetic dissection of complex traits in crops. In this study, phenotypic observations of five silique-related traits replicated in seven environments were used for the QTL analysis, and they might improve the accuracy of QTL detection.

In the present study, most of the five silique-related traits have significant positive correlations with each other (**Table 2**). The significant correlations can be explained by unique pleiotropic QTLs. For example, SB had significant positive correlations with SV (coefficient 0.718), whereas, five unique QTLs simultaneously effect SB and SV with the same direction of additive effects (**Supplementary Additional File 3**). ST showed a negative but low correlations with SL (coefficient -0.026) and SPS (coefficient −0.027) (**Table 4**). Accordingly, no unique QTL had effects on both ST and SL, whereas only two unique QTLs controlled ST

TABLE 3 | Consensus QTL obtained for the five silique-related traits detected in multiple environments.

Trait	QTL	Chr.	Position (cM)	CI. (cM)	LOD	Add	R2%	Env.
SL	*cqSL.A10-3*	A10	34.21	33.66–34.76	2.73–4.07	0.15–0.17	5.25–7.22	14DL/15YL
SB	*cqSB.A10-2*	A10	58.22	56.5–59.9	3.19–5.2	−0.12	6.08–9.04	14DL/16DL/16NJ
	cqSB.C1	C1	11.6	10.7–12.5	2.76–3.42	0.10–0.11	4.61–6.22	15YL/16GS
	cqSB.C6-1	C6	51.25	50.2–52.4	2.5–6.98	0.11–0.15	4.38–13.0	14DL/14NJ/15NJ/16DL
	cqSB.C7-2	C7	92.57	92.16–92.97	4.97–5.79	0.14–0.2	9.22–11.7	15YL/16GS
	cqSB.C7-4	C7	98.52	98.1–99.0	3.91–5.26	0.11–0.14	6.73–9.63	16DL/16GS
ST	*cqST.A5-2*	A5	22.15	21.2–23.11	2.69–4.03	−0.06–−0.07	5.04–7.52	16DL/16GS
	cqST.A5-3	A5	29.04	27.69–30.39	3–3.14	−0.07	5.91–6.16	14DL/14NJ
SPS	*cqSPS.A7-3*	A7	74.09	72.21–75.98	3.61–6.29	0.85–1.26	6.92–11.2	14DL/16DL
	cqSPS.C3-2	C3	33.97	28.09–39.84	2.66–3.17	0.8–0.97	5.49–5.75	15NJ/16GS
	cqSPS.C3-3	C3	41.27	41.08–41.46	3.16–4.47	0.9–0.96	5.76–8.25	15NJ/15YL/16NJ
	cqSPS.C6-2	C6	52.02	50.03–54.01	2.97–3.14	0.78–0.89	5.93–6.13	15YL/16NJ
SV	*cqSV.A10-1*	A10	77.7	77.61–77.8	3.09–4.44	−0.03–−0.04	6.5–8.4	14DL/15NJ

SL, SB, ST, SPS, and SV indicate the traits silique length, silique breadth, silique thickness, seed number per silique, and silique volume, respectively.
Chr., Chromosome.
CI., Confidence interval (cM).
Add, Additive effect.
Env., The experiment in which the consensus QTLs were detected.

and SPS. These were *uqA4-2,* with the same direction of additive effects, and *uqA5-4*, with the opposite direction of additive effects. In addition, SPS had significant positive correlations with SL, SB and SV, but not ST, which is consistent with previous results (Zhang et al., 2011; Wang X. et al., 2016; Yang et al., 2017).

In this study, 25, 29, and 21 consensus QTLs were obtained for SB, ST and SV, respectively. Among these QTLs, five for SB, two for ST and one for SV were repeatedly identified in two to four different environments. To our knowledge, QTLs for SB, ST, and SV in *B. napus* have rarely been reported (Wang X. et al., 2016). In this study, the QTLs for SB were first mapped on chromosomes A1, A10, C1, C3, C4, C8, and C9; QTLs for ST were first mapped on chromosomes A2, A4, C3, and C7; and QTLs for SV were first mapped on chromosomes A2, A4, A7, A8, A10, C2, C3, C4, C7, C8, and C9. Meanwhile, in previously published results, two QTLs related to SB located at 60.47–62.39 Mb and 68.69–72.5 Mb of C6 had overlapping CIs with the results of our study (*cqSB.C6-2*, 60.6–61.9 Mb and *cqSB.C6-3*, 70.1–83.3 Mb), two QTLs related to ST located at 61.3–69.2 Mb of A9 and 34.1–60.4 Mb of A10 had overlapping CIs with the results of our study (*cqST.A9-2*, 68.1–75 Mb and *cqST.A10-1*, 38.9–52 Mb) and one QTL related to SV located at 54.8–60.7 Mb of C6 had an overlapping CI with the results of our study (*cqSV.C6-2*, 50.9–55.9 Mb). This indicated that these QTLs were stably expressed in different genetic backgrounds and that they may be useful in molecular marker-assisted selection breeding.

In previous studies, efforts focused on QTL analyses of SPS and SL. There are more than 200 QTLs for SPS (Radoev et al., 2008; Shi et al., 2009; Chen et al., 2011; Zhang et al., 2011; Cai et al., 2014, 2016; Qi et al., 2014; Wang X. et al., 2016; Yang et al., 2016, 2017; Luo et al., 2017; Zhu et al., 2020) and more than 100 QTLs for SL located on 19 chromosomes (Chen et al., 2007; Zhang et al., 2011; Yang et al., 2012, 2017; Cai et al., 2014; Qi et al., 2014; Fu et al., 2015; Wang X. et al., 2016). Most of the QTLs for SL and SPS obtained in this study were consistent with those previously reported. Among them, one QTL for SL (*cqSL.A10-3*) and four for SPS (*cqSPS.A7-3*, *cqSPS.C3-2*, *cqSPS.C3-3,* and *cqSPS.C6-2*) were repeatedly identified in two to three different environments. With the assistance of the *B. napus* reference genome "Darmor-*bzh*," it was possible to compare the QTLs for silique-related traits detected in this study with those detected in previous studies. *cqSL.A10-3*, *cqSPS.A7-3*, *cqSPS.C3-2*, *cqSPS.C3-3,* and *cqSPS.C6-2* overlapped with QTL mapping intervals of Chen et al. (2007), Shi et al. (2009), Luo et al. (2017), and Zhu et al. (2020), respectively. In this study, the 13 consensus QTLs that were stably expressed in different genetic backgrounds and environments were used as the major QTLs for further candidate gene research. Among these 13 consensus QTLs, *cqSB.C7-4*, *cqST.A5-2*, *cqSPS.A7-3*, *cqSPS.C3-3*, *cqSB.C6-1,* and *cqSPS.C6-2* were further integrated into pleiotropic unique QTLs *uqC7-7*, *uqA5-4*, *uqA7-6*, *uqC3-4,* and *uqC6-4* by meta-analysis (**Table 4**). The genes underlying these consensus QTL might have pleiotropic effects on two to three traits.

Siliques represent a fruit type specific to members of the *Brassicaceae* family that form from the gynoecium after flowering (Seymour et al., 2008). The transformation of the gynoecium into a silique depends on whether the ovule has been successfully fertilized, and this signal may be produced by the pollen grains (Seymour et al., 2013). Successfully fertilized ovules result in developmental alterations of pistils from senescence to growing fruit. Thus, the failure of any part of these successive processes, including gynoecium formation, fertilization and silique growth, that involving cell proliferation, differentiation and expansion, will eventually affect the development of siliques and seeds, which are the determining factors of rapeseed yield. For example, plants with a *BnaC9.SMG7b* deletion exhibit reduced SPS and SL values, which are caused by developmental defects in the formation of functional FGs owing to the interruption of meiotic anaphase II (Li et al., 2015). *BnaC9.SMG7b*, the successfully

TABLE 4 | Twenty-five pleiotropic unique QTL on two to three traits of silique-related traits obtained by the second round of meta-analysis.

Unique QTLs			Consensus QTLs								
QTL	**Position (cM)**	**CI. (cM)**	**QTL**	**Chr.**	**Position (cM)**	**CI. (cM)**	**LOD**	**Add**	**R2%**	**Env.**	**Trait**
uqA1-2	44.05	41.56–46.54	*cqSPS.A1-1*	A1	42.21	41.2–47	2.83	−0.72	5.11	16NJ	SPS
			cqSB.A1-2	A1	49.21	46.2–55.9	3.35	0.11	5.64	16GS	SB
uqA2-3	96.6	93.88–99.32	*cqSV.A2-3*	A2	96.21	92.2–99	2.84	−0.04	5.8	16NJ	SV
			cqST.A2-1	A2	97.31	90.8–99.9	2.53	−0.05	4.63	15YL	ST
uqA4-2	60.9	58.95–62.84	*cqSPS.A4*	A4	57.31	56–62.4	4.02	1.07	7.29	15NJ	SPS
			cqST.A4	A4	63.01	62.3–67.2	2.72	0.06	5.13	16GS	ST
uqA5-4	22.81	21.97–23.64	*cqST.A5-2*	A5	22.15	21.2–23.11	2.69∼4.03	−0.06∼0.71	5.04∼7.52	16DL/16GS	ST
			cqSPS.A5-1	A5	24.91	22.1–25.5	3.02	0.71	5.71	14DL	SPS
uqA7-3	67.53	66.89–68.17	*cqSPS.A7-2*	A7	66.61	64.2–67.4	4.49	1.09	8.44	16DL	SPS
			cqSL.A7-1	A7	67.71	66.6–68	2.93	0.17	5.78	16GS	SL
uqA7-5	70.53	70.35–70.72	*cqSB.A7-2*	A7	70.01	69.7–70.8	5	−0.14	9.12	14DL	SB
			cqSL.A7-2	A7	70.61	70.4–70.8	3.11	0.16	6.03	16GS	SL
uqA7-6	74.81	73.21–76.41	*cqSPS.A7-3*	A7	74.09	72.21–75.98	3.61∼6.29	0.85∼1.26	6.92∼11.2	14DL/16DL	SPS
			cqSB.A7-3	A7	76.71	73.8–79.9	4.02	−0.13	7.79	14DL	SB
uqA9-1	10.03	5.95–14.11	*cqSL.A9*	A9	5.31	3.6–15	2.56	0.15	4.8	16DL	SL
			cqSB.A9	A9	15.01	8.1–19.8	3.77	−0.13	7.1	15NJ	SB
uqA10-4	52.32	51.36–53.28	*cqST.A10-1*	A10	46.31	38.9–52	3.81	−0.069	7.05	16DL	ST
			cqSB.A10-1	A10	52.01	50.4–52.6	5.43	−0.13	10.1	14DL	SB
			cqST.A10-2	A10	54.01	52–56.1	3.75	−0.07	7.07	16DL	ST
uqA10-8	77.7	77.61–77.78	*cqSV.A10-1*	A10	77.7	77.61–77.8	3.09∼4.44	−0.0329∼−0.04	6.5∼8.4	14DL/15NJ	SV
			cqSB.A10-4	A10	77.71	77.6–78.2	2.62	−0.1	4.84	15NJ	SB
			cqST.A10-4	A10	77.71	77.6–78.2	4.39	−0.08	8.34	15NJ	ST
uqA10-10	86.83	86.33–87.33	*cqST.A10-5*	A10	86.51	85.3–87.6	3.42	−0.07	6.43	15NJ	ST
			cqSV.A10-3	A10	86.91	86.6–87.7	4.48	−0.0336	8.62	14DL	SV
uqC2-3	52.45	51.08–53.83	*cqSPS.C2-2*	C2	51.01	49.9–53.8	3.65	0.93	7.14	16GS	SPS
			cqSV.C2	C2	53.91	53.1–57	2.56	0.03	4.98	16GS	SV
uqC3-1	17.2	14.41–20.0	*cqSL.C3-1*	C3	17.21	11.7–23.2	3.34	0.17	6.3	16DL	SL
			cqSPS.C3-1	C3	17.21	13.9–20.7	7.54	1.3	13.8	16DL	SPS
			cqSV.C3	C3	17.21	7.9–27.1	2.6	0.05	4.78	16DL	SV
uqC3-4	41.33	41.15–41.51	*cqSPS.C3-3*	C3	41.27	41.08–41.46	3.16∼4.47	0.9∼0.96	5.76∼8.25	15NJ/15YL/16NJ	SPS
			cqSL.C3-2	C3	42.11	41.4–42.7	2.82	0.16	5.45	15YL	SL
uqC3-6	50.05	48.67–51.43	*cqSPS.C3-5*	C3	50.01	48.1–51	3.31	0.79	6.2	16NJ	SPS
			cqSB.C3	C3	50.51	50.3–59.3	3.46	0.12	6.39	15NJ	SB
uqC6-3	43.21	41.3 - 45.11	*cqSPS.C6-1*	C6	43.21	41.2–46	4.53	0.91	8.31	16NJ	SPS
			cqSV.C6-1	C6	43.21	40.9–47.2	2.85	0.05	5.29	16DL	SV
uqC6-4	51.45	50.55–52.36	*cqSB.C6-1*	C6	51.25	50.14–52.37	2.5∼6.98	0.11∼0.15	4.38∼12.96	14DL/14NJ/15NJ/16DL	SB
			cqSV.C6-2	C6	51.61	50.9–55.9	3.15	0.06	5.89	16DL	SV
			cqSPS.C6-2	C6	52.02	50.03–54.01	2.97∼3.14	0.78∼0.89	5.93∼6.13	15YL/16NJ	SPS

(Continued)

TABLE 4 | (Continued)

Unique QTLs			Consensus QTLs								
QTL	**Position (cM)**	**CI. (cM)**	**QTL**	**Chr.**	**Position (cM)**	**CI. (cM)**	**LOD**	**Add**	**R2%**	**Env.**	**Trait**
uqC7-7	98.59	98.21–98.96	*cqSV.C7-1*	C7	98.31	97.1–100.3	4.15	0.07	7.77	16DL	SV
			cqSB.C7-4	C7	98.52	98.06–98.97	3.91∼5.26	0.11∼0.14	6.73∼9.63	16DL/16GS	SB
			cqST.C7-1	C7	98.81	97.6–99	4.13	0.07	7.79	16DL	ST
uqC7-8	99.2	99.12–99.29	*cqSB.C7-5*	C7	99.21	99.1–99.3	7.25	0.2	14.16	15YL	SB
			cqSV.C7-2	C7	99.21	99–99.3	5.15	0.04	10.3	16GS	SV
uqC7-9	101.51	101.18–101.83	*cqSB.C7-6*	C7	101.51	101.2–101.9	5.81	0.15	10.58	16GS	SB
			cqST.C7-2	C7	101.51	101.2–103.1	2.76	0.08	5.42	14NJ	ST
uqC7-10	103.51	103.3–105.0	*cqST.C7-3*	C7	103.51	103.3–105	3.11	0.07	5.85	16GS	ST
			cqSB.C7-7	C7	104.11	103.3–104.5	7.54	0.22	14.63	15YL	SB
uqC7-13	117.89	116.96–118.82	*cqST.C7-5*	C7	117.41	116–118	3.23	−0.1	6.2	14NJ	ST
			cqSV.C7-4	C7	120.91	117.5–122.5	3.99	0.031	7.54	14DL	SV
uqC8	14.19	13.71–14.67	*cqSB.C8*	C8	14.11	13.4–14.4	3.45	0.12	6.37	15NJ	SB
			cqSV.C8	C8	15.01	14.8–24.6	2.62	0.03	4.98	15NJ	SV
			cqSPS.C8	C8	15.11	14.4–17.9	3.77	1.01	6.84	15NJ	SPS
uqC9-5	115.47	114.65–116.28	*cqST.C9-4*	C9	115.11	113–115.6	3.54	0.065	6.56	15YL	ST
			cqSB.C9	C9	115.71	114.2–116.3	5.25	0.14	9.96	15YL	SB
uqC9-6	117.01	116.02–117.99	*cqST.C9-5*	C9	117.01	116.4–118.8	3.27	0.08	5.99	16NJ	ST
			cqSV.C9	C9	117.01	116.3–119.8	2.54	0.03	4.72	15YL	SV

SL, SB, ST, SPS, and SV indicate the traits silique length, silique breadth, silique thickness, seed number per silique, and silique volume, respectively.
DL, Dali; NJ, Nanjing; YL, Yangling; GS, Gansu; 14, 15, and 16 indicate the years 2014, 2015 and 2016, respectively.
Chr., Chromosome.
CI., Confidence interval (cM).
Add, Additive effect.
Env., The experiment in which the consensus QTLs were detected.

cloned rapeseed silique gene, also underlies the CI of *cqSL.A10-3* in this study. *LNG1* regulates longitudinal cell elongation in *Arabidopsis*, and *lng1* dominant mutant plants are characterized by elongated siliques owing to longitudinally elongated and transversely narrowed cells in the siliques (Lee et al., 2006). CNGC18 is an essential Ca^{2+} channel for pollen tube guidance in *Arabidopsis*. *CNGC18* point mutations result in shorter siliques, reduced male fertility and fewer seeds per silique (Gao Q. et al., 2016). The *Arabidopsis* gene *STY2* promotes the formation of the apical tissues of the gynoecium. The *sty2* mutant lines exhibit many developmental defects in reproductive tissues, including shortened siliques and aborted ovules (Kuusk et al., 2002). *KIN*βγ, a component of the regulatory subunit of the *SNF1*-related protein kinase, is required for pollen germination on the stigma surface. *KIN*βγ mutants display defects in organogenesis and growth, including shorter statures and siliques (Gao X. et al., 2016). *ABCC13* is expressed in the seed coat and embryo, and its loss of function in *Arabidopsis* leads to decreased silique lengths and seed yields (El Guizani et al., 2014). *ARID1* is required for sperm cell formation in *Arabidopsis*, and the *ARID1* mutant shows reduced seed set and short siliques, which are caused by defects in gametophyte formation owing to an arrested mitotic cell cycle (Zheng et al., 2014). *LAC4*, a laccase gene regulated by miR397b, controls both lignin biosynthesis and seed yield in *Arabidopsis*, and overexpressing miR397b may increase silique lengths and seed sizes by modulating *LAC4* (Wang et al., 2014). *HTH* is involved in regulating floral organ fusion in *Arabidopsis*, and the *hth* mutant shows reduced seed set and short siliques, which are caused by reduced pollen fertility and aborted ovules (Krolikowski et al., 2003). The homozygous *aog1* in *Arabidopsis* shows reduced seed set and short siliques compared with the wild type, and this caused by a significant reduction in fertility owing to reduced pollen formation and severe defects in embryo sacs (Cui et al., 2015). PYL8 is a regulatory component of the ABA receptor, which is vital for regulating seed germination, root and shoot development and abiotic stress responses (Garcia-Maquilon et al., 2021), and a *pyl8* mutant shows an ABA-insensitive phenotype, including a lower seed yield owing to fewer siliques having shorter lengths (Gonzalez-Guzman et al., 2012). *Arabidopsis HMG1* encodes a 3-hydroxy-3-methylglutaryl coenzyme A reductase, which is a key element of the sterol biosynthetic process that is required for cell viability and growth. The T-DNA insertion *hmg1* mutant in *Arabidopsis* shows shorter siliques and fewer seeds per silique compared with the wild type, and this was caused by reduced cell elongation and fertility owing to a reduced sterol level (Suzuki et al., 2004).

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/s.

AUTHOR CONTRIBUTIONS

XZ and KY co-wrote the manuscript. CS, WZ, and FC participated in the field experiment and collected the data. CP, XWu, and RS carried out QTL analysis and prediction of candidate genes. JZ revised the manuscript. XWa designed, led and coordinated the overall study. All authors have read and approved the final manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (31971973 and 32060495), Jiangsu Agricultural Science and Technology Innovation Fund [CX(19)3053], China Agriculture Research System of MOF and MARA (CARS-12), and Jiangsu Collaborative Innovation Center for Modern Crop Production.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.766271/full#supplementary-material

Supplementary Additional File 1 | The 137 identified QTLs and 120 consensus QTLs for five silique-related traits detected in seven environments.

Supplementary Additional File 2 | Consensus QTL locations of silique-related traits detected from seven different environments.

Supplementary Additional File 3 | The 89 unique QTLs associated with five silique-related traits in the AH population.

Supplementary Additional File 4 | Genes underlying the confidence intervals of the 13 stable QTLs.

REFERENCES

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095

Allen, E. J., Morgan, D. G., and Ridgman, W. J. (1971). A physiological analysis of the growth of oilseed rape. *J. Agric. Sci.* 77, 339–341. doi: 10.1017/S0021859600024515

Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A., et al. (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 20, 2324–2326. doi: 10.1093/bioinformatics/bth230

Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., et al. (2014). Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 127, 85–96. doi: 10.1007/s00122-013-2203-9

Cai, G., Yang, Q., Chen, H., Yang, Q., Zhang, C., Fan, C., et al. (2016). Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Rep.* 6:21625. doi: 10.1038/srep21625

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

Chen, F., Zhang, W., Yu, K., Sun, L., Gao, J., Zhou, X., et al. (2018). Unconditional and conditional QTL analyses of seed fatty acid composition in *Brassica napus* L. *BMC Plant Biol.* 18:49. doi: 10.1186/s12870-018-1268-7

Chen, W., Zhang, Y., Liu, X., Chen, B., Tu, J., and Tingdong, F. (2007). Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F_2 populations. *Theor. Appl. Genet.* 115, 849–858. doi: 10.1007/s00122-007-0613-2

Chen, W., Zhang, Y., Yao, J., Ma, C., Tu, J., and Tingdong, F. (2011). Quantitative trait loci mapping for two seed yield component traits in an oilseed rape (*Brassica napus*) cross. *Plant Breed.* 130, 640–646. doi: 10.1111/j.1439-0523.2011.01886.x

Clarke, J. M., and Simpson, G. M. (1978). Influence of irrigation and seeding rates on yield and yield components of *Brassica napus* cv. tower. *Can. J. Plant Sci.* 58, 731–737. doi: 10.4141/cjps78-108

Cui, H., Liao, H., Tang, Y., Du, X., Chen, L., Ye, D., et al. (2015). ABORTED GAMETOPHYTE 1 is required for gametogenesis in *Arabidopsis*. *J. Integr. Plant Biol.* 57, 1003–1016. doi: 10.1111/jipb.12341

Danecek, P., and McCarthy, S. A. (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33, 2037–2039. doi: 10.1093/bioinformatics/btx100

Dong, Z., Alam, M. K., Xie, M., Yang, L., Liu, J., Helal, M., et al. (2021). 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 (Bethesda)* doi: 10.1093/g3journal/jkab118 [Epub Online ahead of print].

El Guizani, T., Blanc, N., Triki, S., St-Pierre, B., and Ducos, E. (2014). Expression pattern of AtABCC13/MRP11 reveals developmental, hormonal, and nutritional regulations. *Biol. Plant.* 58, 231–240. doi: 10.1007/s10535-013-0387-0

Ferrándiz, C., Pelaz, S., and Yanofsky, M. F. (1999). Control of carpel and fruit development in *Arabidopsis*. *Annu. Rev. Biochem.* 68, 321–354. doi: 10.1146/annurev.biochem.68.1.321

Fu, Y., Wei, D., Dong, H., He, Y., Cui, Y., Mei, J., et al. (2015). Comparative quantitative trait loci for silique length and seed weight in *Brassica napus*. *Sci. Rep.* 5:14407. doi: 10.1038/srep14407

Gao, Q., Gu, L., Wang, H., Fei, C., Fang, X., Hussain, J., et al. (2016). Cyclic nucleotide-gated channel 18 is an essential Ca^{2+} channel in pollen tube tips for pollen tube guidance to ovules in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3096–3101. doi: 10.1073/pnas.1524629113

Gao, X., Liu, C., Li, D., Zhao, T., Li, F., Jia, X., et al. (2016). The *Arabidopsis* KINβγ subunit of the SnRK1 complex regulates pollen hydration on the stigma by mediating the level of reactive oxygen species in pollen. *PLoS Genet.* 12:e1006228. doi: 10.1371/journal.pgen.1006228

Garcia-Maquilon, I., Coego, A., Lozano-Juste, J., Messerer, M., de Ollas, C., Julian, J., et al. (2021). PYL8 ABA receptors of Phoenix dactylifera play a crucial role in response to abiotic stress and are stabilized by ABA. *J. Exp. Bot.* 72, 757–774. doi: 10.1093/jxb/eraa476

Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., et al. (2012). *Arabidopsis* PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. *Plant Cell* 24, 2483–2496. doi: 10.1105/tpc.112.098574

Krolikowski, K. A., Victor, J. L., Wagler, T. N., Lolle, S. J., and Pruitt, R. E. (2003). Isolation and characterization of the *Arabidopsis* organ fusion gene HOTHEAD. *Plant J.* 35, 501–511. doi: 10.1046/j.1365-313X.2003.01824.x

Kuusk, S., Sohlberg, J. J., Long, J., Fridborg, I., and Sundberg, E. (2002). STY1 and STY2 promote the formation of apical tissues during *Arabidopsis* gynoecium development. *Development* 129, 4707–4717. doi: 10.1242/dev.129.20.4707

Lee, Y. K., Kim, G. T., Kim, I. J., Park, J., Kwak, S. S., Choi, G., et al. (2006). LONGIFOLIA1 and LONGIFOLIA2, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. *Development* 133, 4305–4314. doi: 10.1242/dev.02604

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.

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

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. U. S. A.* 112, E5123–32. doi: 10.1073/pnas.1502160112

Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., et al. (2017). Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585. doi: 10.1007/s00122-017-2911-7

McCouch, S. R., Cho, Y. G., Yano, M., Paul, E., Blinstrub, M., Morishima, H., et al. (1997). Report on QTL nomenclature. *Rice Genet. Newsl.* 14, 11–13.

Merk, H. L., Yarnes, S. C., Van Deynze, A., Tong, N., Menda, N., Mueller, L. A., et al. (2012). Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato germplasm. *J. Am. Soc. Hortic. Sci.* 137, 427–437. doi: 10.21273/JASHS.137.6.427

Murphy, D. J. (1999). "The future of new and genetically modified oil crops," in *Perspectives on New Crops and New Uses*, ed. J. Janick (Alexandria: ASHS Press), 216–219.

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

Radoev, M., Becker, H. C., and Ecke, W. (2008). Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by QTL mapping. *Genetics* 179, 1547–58. doi: 10.1534/genetics.108.089680

Riehs, N., Akimcheva, S., Puizina, J., Bulankova, P., Idol, R. A., Siroky, J., et al. (2008). *Arabidopsis SMG7* protein is required for exit from meiosis. *J. Cell Sci.* 121, 2208–2216. doi: 10.1242/jcs.027862

Seymour, G., Poole, M., Manning, K., and King, G. J. (2008). Genetics and epigenetics of fruit development and ripening. *Curr. Opin. Plant Biol.* 11, 58–63. doi: 10.1016/j.pbi.2007.09.003

Seymour, G. B., Østergaard, L., Chapman, N. H., Knapp, S., and Martin, C. (2013). Fruit development and ripening. *Annu. Rev. Plant Biol.* 64, 219–241. doi: 10.1146/annurev-arplant-050312-120057

Shi, J., Li, R., Qiu, D., Jiang, C., Long, Y., Morgan, C., et al. (2009). Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics* 182, 851–861. doi: 10.1534/genetics.109.101642

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

Sotelo-Silveira, M., Cucinotta, M., Chauvin, A. L., Chávez, M. R., Colombo, L., Marsch-Martínez, N., et al. (2013). Cytochrome P450 *CYP78A9* is involved in *Arabidopsis* reproductive development. *Plant Physiol.* 162, 779–799. doi: 10.1104/pp.113.218214

Sun, C., Wang, B., Yan, L., Hu, K., Liu, S., Zhou, Y., et al. (2016). Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 7:1102. doi: 10.3389/fpls.2016.01102

Sun, L., Wang, X., Yu, K., Li, W., Peng, Q., Chen, F., et al. (2018). Mapping of QTLs controlling seed weight and seed-shape traits in *Brassica napus* L. using a high-density SNP map. *Euphytica* 214:228. doi: 10.1007/s10681-018-2303-3

Suzuki, M., Kamide, Y., Nagata, N., Seki, H., Ohyama, K., Kato, H., et al. (2004). Loss of function of 3-hydroxy-3-methylglutaryl coenzyme a reductase 1 (*HMG1*) in *Arabidopsis* leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.* 37, 750–761. doi: 10.1111/j.1365-313x.2004.02003.x

Van Camp, W. (2005). Yield enhancement genes: seeds for growth. *Curr. Opin. Biotech.* 16, 147–153. doi: 10.1016/j.copbio.2005.03.002

Wang, C., Zhang, S., Yu, Y., Luo, Y., Liu, Q., Ju, C., et al. (2014). MiR397b regulates both lignin content and seed number in *Arabidopsis via* modulating a laccase involved in lignin biosynthesis. *Plant Biotechnol. J.* 12, 1132–1142. doi: 10.1111/pbi.12222

Wang, H. (2010). Review and future development of rapeseed industry in China (in Chinese with an English abstract). *Chin. J. Oil Crop Sci.* 32, 300–302. doi: 10.3724/SP.J.1011.2010.01385

Wang, H., Cheng, H., Wang, W., Liu, J., Hao, M., Mei, D., et al. (2016). Identification of *BnaYUCCA6* as a candidate gene for branch angle in *Brassica napus* by QTL-seq. *Sci. Rep.* 6:38493. doi: 10.1038/srep38493

Wang, S., Basten, C., and Zeng, Z. (2012). *Windows QTL Cartographer 2.5*. Raleigh: North Carolina State University.

Wang, W., Chu, W., Mei, D., Cheng, H., Zhu, L., Fu, L., et al. (2019). Quantitative trait loci mapping for branch angle and candidate gene screening in *Brassica napus* L (in Chinese with an English abstract). *Acta Agronom. Sin.* 45, 37–45. doi: 10.3724/SP.J.1006.2019.84042

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. doi: 10.1186/s12870-016-0759-7

Wang, X., Yu, K., Li, H., Peng, Q., Chen, F., Zhang, W., et al. (2015). High-density SNP map construction and QTL identification for the apetalous character in *Brassica napus* L. *Front. Plant Sci.* 6:1164. doi: 10.3389/fpls.2015.01164

Wang, Y., Chen, W., Chu, P., Wan, S., Yang, M., Wang, M., et al. (2016). Mapping a major QTL responsible for dwarf architecture in *Brassica napus* using a single-nucleotide polymorphism marker approach. *BMC Plant Biol.* 16:178. doi: 10.1186/s12870-016-0865-6

Xu, Y., Zhang, B., Ma, N., Liu, X., Qin, M., Zhang, Y., et al. (2021). Quantitative trait locus mapping and identification of candidate genes controlling flowering time in *Brassica napus* L. *Front. Plant Sci.* 11:626205. doi: 10.3389/fpls.2020.626205

Yang, P., Shu, C., Chen, L., Xu, J., Wu, 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

Yang, Y., Shen, Y., Li, S., Ge, X., and Li, Z. (2017). High density linkage map construction and QTL detection for three silique-related traits in *Orychophragmus violaceus* derived *Brassica napus* population. *Front. Plant Sci.* 8:1512. doi: 10.3389/fpls.2017.01512

Yang, Y., Shi, J., Wang, X., Liu, G., and Wang, H. (2016). Genetic architecture and mechanism of seed number per pod in rapeseed: elucidated through linkage and near-isogenic line analysis. *Sci. Rep.* 6:24124. doi: 10.1038/srep24124

Yu, K., Wang, X., Li, W., Sun, L., Peng, Q., Chen, F., et al. (2019). Identification and physical mapping of QTLs associated with flowering time in *Brassica napus* L. *Euphytica* 215:152. doi: 10.1007/s10681-019-2480-8

Yu, K., Zhang, W., Guo, Y., Zheng, M., Chen, F., Sun, C., et al. (2021). Integrating unconditional and conditional QTLs to dissect the genetic basis of stem mechanical strength in *Brassica napus* L. *Euphytica* 217:34. doi: 10.1007/s10681-021-02769-0

Zhang, L., Li, S., Chen, L., and Yang, G. (2012). Identification and mapping of a major dominant quantitative trait locus controlling seeds per silique as a single Mendelian factor in *Brassica napus* L. *Theor. Appl. Genet.* 125, 695–705. doi: 10.1007/s00122-012-1861-3

Zhang, L., Yang, G., Liu, P., Hong, D., Li, S., and He, Q. (2011). Genetic and correlation analysis of silique-traits in *Brassica napus* L. by quantitative trait locus mapping. *Theor. Appl. Genet.* 122, 21–31. doi: 10.1007/s00122-010-1419-1

Zheng, B., He, H., Zheng, Y., Wu, W., and McCormick, S. (2014). An ARID domain-containing protein within nuclear bodies is required for sperm cell formation in *Arabidopsis thaliana*. *PLoS Genet.* 10:e1004421. doi: 10.1371/journal.pgen.1004421

Zhu, Y., Ye, J., Zhan, J., Zheng, X., Zhang, J., Shi, J., et al. (2020). Validation and characterization of a seed number per silique quantitative trait locus *qSN.A7* in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 11:68. doi: 10.3389/fpls.2020.00068

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 13 September 2022
DOI 10.3389/fpls.2022.999790

OPEN ACCESS

EDITED BY
Kun Lu,
Southwest University, China

REVIEWED BY
Liezhao Liu,
Southwest University, China
Lunwen Qian,
Hunan Agricultural University, China
Benbo Xu,
Yangtze University, China

*CORRESPONDENCE
Jiaqin Shi
shijiaqin@caas.cn
Xinfa Wang
wangxinfa@caas.cn

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

RECEIVED 21 July 2022
ACCEPTED 16 August 2022
PUBLISHED 13 September 2022

CITATION
Qadir M, Qin L, Ye J, Ahmad N, Wang X, Shi J and Wang H (2022) Genetic dissection of the natural variation of ovule number per ovary in oilseed rape germplasm (*Brassica napus* L.). *Front. Plant Sci.* 13:999790. doi: 10.3389/fpls.2022.999790

Genetic dissection of the natural variation of ovule number per ovary in oilseed rape germplasm (*Brassica napus* L.)

Muslim Qadir[1], Lei Qin[1], Jiang Ye[1], Nazir Ahmad[1], Xinfa Wang[1,2]*, Jiaqin Shi[1]* and Hanzhong Wang[1,2]

[1]Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Wuhan, China, [2]Hubei Hongshan Laboratory, Wuhan, China

Oilseed rape is one of the world's largest oil and industrial crops, providing humans with various products, such as vegetable oil and biofuel. Ovules are the direct precursors of seeds, and ovule number per ovary (ONPO) largely determines seed number per fruit that affects both yield and fitness of seed crops. The ONPO shows wide variation in oilseed rape, whereas the underlying genes and mechanisms are poorly known. The present study performed the genetic, physiological and transcriptomic analyses of ovule number per ovary using an association panel and the extreme lines. The ONPO of 327 accessions planted in four environments showed a large variation from 19.2 to 43.8, indicating a great potential for the further genetic improvement of ovule number. The genome-wide association study (GWAS) identified a total of 43 significant SNP markers. Further, these SNPs were integrated into 18 association loci, which were distributed on chromosomes A01, A03, A06, A07, A09, C01, C03, C06, C07, and C09, explaining 4.3–11.5% of the phenotypic variance. The ONPO decreased as their appearance order on the inflorescence and was associated with the level of several types of endogenous phytohormones but not related to leaf area and photosynthetic rate. Comparative transcriptomic analysis identified a total of 4,449 DEGs enriched in 30 classes, including DNA, RNA, protein, signaling, transport, development, cell wall, lipid metabolism, and secondary metabolism. Nearly half of DEGs were involved in the known pathways in regulating ovule number, of which 12 were homologous to know ovule number regulating genes, indicating a strong link between the identified DEGs and ovule number. A total of 73 DEGs were located within the genomic regions of association loci, of which six were identified as candidates based on functional annotation. These results provide useful information for the further genetic improvement of ovule and seed number in oilseed rape.

KEYWORDS

ovule number per ovary, genome-wide association study, transcriptomic analysis, QTLs, candidate genes, phytohormones, *Brassica napus*

Introduction

Brassica napus L. (AACC, 2*n* = 38) is an allopolyploid (AACC) species generated from an interspecies crossing between *B. rapa* (AA, 2*n* = 20) and *B. oleracea* (CC, 2*n* = 18) around 7,500 years ago (Chalhoub et al., 2014; Wang et al., 2020; Bilgrami et al., 2022). Oilseed rape is one of the most important oilseeds crops worldwide after soybean that provides high-quality nutrients and nutraceuticals to humans and animals and biofuel for industrial production (Yang et al., 2017; Khan et al., 2019). With the increasing demand for edible oil and biofuel due to population growth, it is an urgent need to improve seed yield per unit area in *B. napus* (Hu Q. et al., 2017; Ahmad et al., 2021; Li et al., 2021).

Under the same planting area, seed yield per unit area is determined by seed yield per plant. In *B. napus*, seed yield per plant is a complex quantitative trait that is determined by three components: silique number per plant, seed number per silique, and seed weight in oilseed rape (Clarke and Simpson, 1978; Li et al., 2015). Of these, the seed number per silique relies on the ovule number per ovary (ONPO), the proportion of ovules to be fertilized and the proportion of fertilized ovules to develop into seeds (Yang et al., 2017; Yuan and Kessler, 2019). Ovules are the direct precursors of seeds containing the female gametophytes, which are fecundated during pollination to inaugurate seed development (Yuan and Kessler, 2019). Therefore, increasing the number of ovules per flower has become an important strategy for improving seed crop yield and addressing food security (Jiao et al., 2021). The number of ovules per flower varies considerably across different species and even among the different accessions of the same species (Burd et al., 2009; Khan et al., 2019). Ovule number per ovary is determined by the ovule initiation process and is significantly affected by flower size and position (Gomez et al., 2018; Yuan and Kessler, 2019), nutrient availability (Strelin and Aizen, 2018), and phytohormones levels (Barro-Trastoy et al., 2020; Qadir et al., 2021). Although much research into seed number per silique has been reported in *B. napus* (Yang et al., 2017; Khan et al., 2019; Zhu et al., 2020), few were performed on ovule number per ovary (Khan et al., 2019; Jiao et al., 2021). Therefore, little is known about the natural variation of ovule number per ovary in oilseed rape germplasm.

Although none of the ovule number QTLs in *Brassica* has been cloned, nearly one hundred ovule number regulating genes have been reported in plants (mainly from *Arabidopsis*), which can be used as the reference for *Brassica*, due to their close relationship (Qadir et al., 2021). The systematic summarization of these ovule number genes showed that it is governed by an integrated genetic and phytohormones network where AUX, BR, and CK are the positive regulator of ovule number, whereas GA acts negatively on it (Nemhauser et al., 2000; Bencivenga et al., 2012; Yuan and Kessler, 2019). For example, *PIN1* is one of the eight transmembrane auxin transporters in *Arabidopsis* which is involved in polar auxin transport and required for ovule primordia formation (Benková et al., 2003; Galbiati et al., 2013). The *pin1* mutant shows reduced auxin transport activity and multiple growth and development defects, including reduced ovule number (Okada et al., 1991).

Genome-wide association analysis can detect the causal loci underlying complex quantitative traits at the whole-genome level, which may contain several to hundreds of genes dependent on the LD decay at these loci. Transcriptomic analysis can quantify gene expression level and identify the differentially expressed gene in the given tissues at the specific stage, which may be upstream causal genes or downstream target genes in regulating trait variation. Numerous studies have demonstrated that the integration of genome-wide association and transcriptomic analysis has become an efficient strategy for identifying candidate genes underlying complex quantitative agronomic traits (Wang et al., 2020; Helal et al., 2021). For example, by GWAS of silique length and transcriptomic analysis of silique wall at 15-DAF, *BnaA9.CYP78A9* was successfully identified as the causal gene for a major association loci *qSL.A09-3* (Hussain et al., 2021). The present study aims to dissect the genetic, physiological, and molecular basis for the natural variation of ovule number per ovary in oilseed rape (*Brassica napus* L.). These results will provide a solid basis for further gene cloning and genetic improvement of ovule number in oilseed rape.

Materials and methods

Plant materials and field trials

The association population was composed of 327 oilseed rape accessions (Li et al., 2020). The field experiments were conducted in four environments, including May 2020 at Ping'an district (36.47°N, 102.09°E) in Haidong city of Qinghai province, Oct 2020 at Yangluo (31.84°N, 114.8°E), and Oct 2021 at both Wuchang (30.35°N, 114.33°E) and Yangluo district in Wuhan city of Hubei province. The field planting followed a randomized complete block design with three biological replications. The field management was conducted according to the local standard practices.

Investigation of ovule number

At the beginning of flowering, three buds were collected from the bottom of the main inflorescence of five plants randomly, resulting in 15 buds for each replication. In total, 58,860 buds (4 environments × 327 lines × 3 replications × 15 buds) were sampled. The sampled buds were fixed in FAA, of which 10 randomly selected buds were dissected for further observation. The calyx and petals of sampled bud were

removed by a dissecting needle and taken out of the ovary. The ovaries were kept in 2 ml Eppendorf tubes containing 90% alcohol solution twice and then washed with ddH_2O. The trichloroacetaldehyde hydrate solutions were added in a small amount to submerge the sample and make the ovaries transparent. The ovaries were kept in a transparent solution for 12 h to 3 days' maximum and then transferred to glass slides under the microscope (SZX2-ILLT, Olympus Corporation, Japan). The ovule numbers were counted manually following previously described methods (Yang et al., 2017; Ali, 2018; Yu et al., 2020).

Observation of ovule number per ovary in different positions

To investigate whether ovule number varied with the physical position of buds on the inflorescence. The ovule number variation from the bottom to the top of the inflorescence was investigated using two representative extreme lines (Yang et al., 2017). Briefly, the ovule number data was investigated from five consecutive buds (such as 1–5, 6–10) before flowering. The buds from the bottom to the top of the main inflorescence of the representative plants were sampled.

Detection of phytohormones

To investigate whether the ovule number variation is associated with the concentrations of endogenous phytohormones, the ovaries at the ovule initiation stage were measured for extreme lines. The buds of 0.5–1 mm length were dissected by hand under a stereomicroscope within the required temperature (Vera-Sirera et al., 2016; Gomez et al., 2018), and the obtained ovaries were mixed equally to generate two pools with three biological replications. The quantification of specified hormones including; Abscisic Acid (ABA), Benzyl Adenine (BA), Gibberellic Acid (GA4), Indole-3-Acetic Acid (IAA), Indole-3-Butyric Acid (IBA), Jasmonic Acid (JA) and Salicylic Acid (SA) were carried out on the phytohormones platform at Huazhong Agricultural University, Wuhan, China.

Measurement of leaf area and photosynthetic rate

At the early flowering time, ten mature leaves were collected from five representative plants of the extreme lines and transferred to the laboratory. The procedure for leaf area measurement was described as in the previous studies (Hu M. et al., 2017; Li et al., 2020). A portable photosynthesis system (LI-6800XT, LI-COR) was used to measure the photosynthetic parameter of the extreme lines in the field between 9:30–11:00 and 14:30–16:00 (Li et al., 2019; Wang et al., 2020).

Genome-wide association analysis

The *Brassica* 50 K SNP array[1] was used for the genotypic analysis of the association population, which contains 45,707 SNP markers. The parameters were set as a missing rate $\leq$0.2, heterozygous rate $\leq$0.2, and minor allele frequency $>$0.05 to examine SNP data using Illumina Bead Studio genotyping software.[2] The probe sequences of these SNP markers were compared to the *B. napus* Darmor-bzh reference genome to identify their physical positions (Chalhoub et al., 2014). The four statistical models were used for the association study, including the general linear model (GLM) with controlling for population structure (Q) and principal component analysis (PCA), the mixed linear model (MLM) controlling for both Q and PCA with relative kinship (K) according to Liu et al. (2016). The GWAS analyses were performed using TASSEL v.5.2.77 software. The threshold for the significantly associated SNP markers was set to $P < 4.08 \times 10^{-5}$ [P = 1/21242, $-\log_{10}(P)$ = 4.33] as previously described (Liu et al., 2016; Li et al., 2020).

Ribonucleic acid sequencing and transcriptomic analysis

The ovaries at the ovule initiation stage were dissected from 0.5 to 1 mm buds by hand under a stereomicroscope at the required temperature. The obtained ovaries were mixed equally to generate two pools with three biological replications (M1-M3 and L1-L3). According to the manufacturer's procedure, a Plant RNA Mini Kit was used to isolate total RNA from each sample (Tiangen, Inc., China). The Oebiotech company performed cDNA library construction and Illumìna sequencing using an Illumìna HìSeq^{TM} 2,500 platform.

The low-quality, low-complexity, and repetitive raw reads were sorted out, and only the clean reads that passed quality control were subjected to further analysis. The sequences were mapped to the reference genome of Darmor as previously described (Hussain et al., 2021). The false discovery rate (FDR) $\leq$0.05) and the p-value $\leq$0.005 were set as a threshold to identify the deferentially expressed genes (DEGs). Moreover, these paired-end sequencing reads were uploaded to NCBI with accession number PRJNA820145.

1 https://www.greenfafa.com/

2 http://www.Illumina.com/

Validation of deferentially expressed genes through qRT-PCR

Ten DEGs randomly chosen were subjected to qRT-PCR validation using the same RNA samples used for sequencing. The M-MLV reverse transcriptase (Promega) was used to synthesize cDNA containing total RNA (4 μg) and oligo (dT) primers, as described in the company instructions. Bio-Rad CFX96 real-time detection systems were used for qRT-PCR analyses with three replicates. The $2^{-\Delta\Delta Ct}$ method was adopted to evaluate the relative expression of target genes, with the *B. napus* ACTIN2 as an internal control (Li et al., 2021). The details of these gene-specific primers used for qRT-PCR are listed in **Supplementary Table 1**.

Identification of candidate genes

Based on the physical distance of LD decay, the genes within 500 kb up and downstream of significantly associated SNP loci were considered (Li et al., 2020). Some of these genes were homologous to the known ovule number regulating genes in plants (Qadir et al., 2021), which were identified as candidates. In addition, some of these genes showed significantly differential expression between the more and less-ovule lines, which were also identified as candidates.

Statistical analysis

The correlation coefficient was calculated using the CORR procedure implemented in the SAS software 8.0 (SAS Institute, Inc., 2000, Cary, NC, United States). Broad-sense heritability was calculated as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$, where σ_g^2, σ_{ge}^2, and σ_e^2 are the variances of genotype, genotype by environment and error, respectively, while n and r are the number of environments and replicates, respectively. The Excel statistical functions CHISQ.TEST and T.TEST were used to obtain the significance level ($P_x^2{}_{-test}$ and P_{t-test}) of the degree-of-fit and differences (Meng et al., 2015). The frequency distributions of ONPO investigated in four environments were constructed using Minitab 9.1 software.

Results

Phenotypic variation of ovule number per ovary

The ONPO of two representative lines (3S1305 and 3S1195) were investigated from every five buds sampled from bottom to top of the main inflorescence. The results showed that the calculated ONPO of the two lines ranged from 40.0 to 31.0 and 23.5 to 16.6, respectively (**Figure 1A**), showing an obvious continuous downward trend in order of occurrence/differentiation. Therefore, the ONPO of the association population was investigated from the buds on the bottom of the main inflorescence.

The descriptive statistics of ONPO of the association population in four environments were presented in **Table 1**. In general, the ONPO of 327 lines in the association panel varied from 19.2 to 43.8 across the four investigated environments, displaying more than two-fold variation. The coefficient of variation of ONPO in the four environments was 8.91, 12.64, 11.64, and 11.50%, respectively (**Table 1**). In general, the high correlations of ONPO were observed between pair-wise environments (r^2 = 0.71, 0.79 and 0.87) except for 2020QH (r^2 = 0.56, 0.60 and 0.63). The phenotypic frequency distribution analysis showed that the ONPO was almost normally distributed in each environment, indicating a quantitative inheritance suitable for QTL mapping (**Figure 1B**). Furthermore, the analysis of variance showed that the variation of ONPO in this association panel was primarily attributed to the genotypes (**Supplementary Table 2**). The calculated broad-sense heritability from variance components was 91.9%, indicating the high stability of this trait suitable for genetic study.

Genome-wide association study of ovule number per ovary

In total, 21,242 SNPs data previously generated in our group were used for the association study (**Figure 2** and **Supplementary Figure 1**). A total of 43 SNP loci were significantly associated with the ONPO (**Supplementary Table 3**). Specifically, 28, 1, 6, and 7 loci were detected in 20QH, 21WC, 20YL, and 21YL, respectively. It should be noted that many SNP loci were detected by multiple models and environments, suggesting their reliability. After the integration of close SNP loci within 500 kb, a total of 18 QTLs were obtained, which were distributed on chromosomes A01, A03, A06, A07, A09, C01, C03, C06, C07, and C09, explaining 4.3 to 11.5% of the phenotypic variance (**Table 2**).

It should be noted that two associated loci (*qONPO.A03-2* and *qONPO.A06-1*) were repeatedly detected in 20YL and 21YL environments, and one (*qONPO.C03-2*) was repeatedly identified in 20QH and 21YL environments. Therefore, they may represent important targets for marker-assisted selection and further gene cloning. For *qONPO.A03-2*, the two variant bases A and G of peak SNP seq-new-rs32414, respectively accounted for 75.5 and 24.5%. The G allele's mean ONPO (30.7) was significantly greater than that (29.0) of the A allele with $P = 2.30E^{-4}$ (**Figure 3**). For *qONPO.A06-1*, the two variant bases A and C of peak SNP seq-new-rs23621, respectively, accounted for 25.2 and 74.8%.

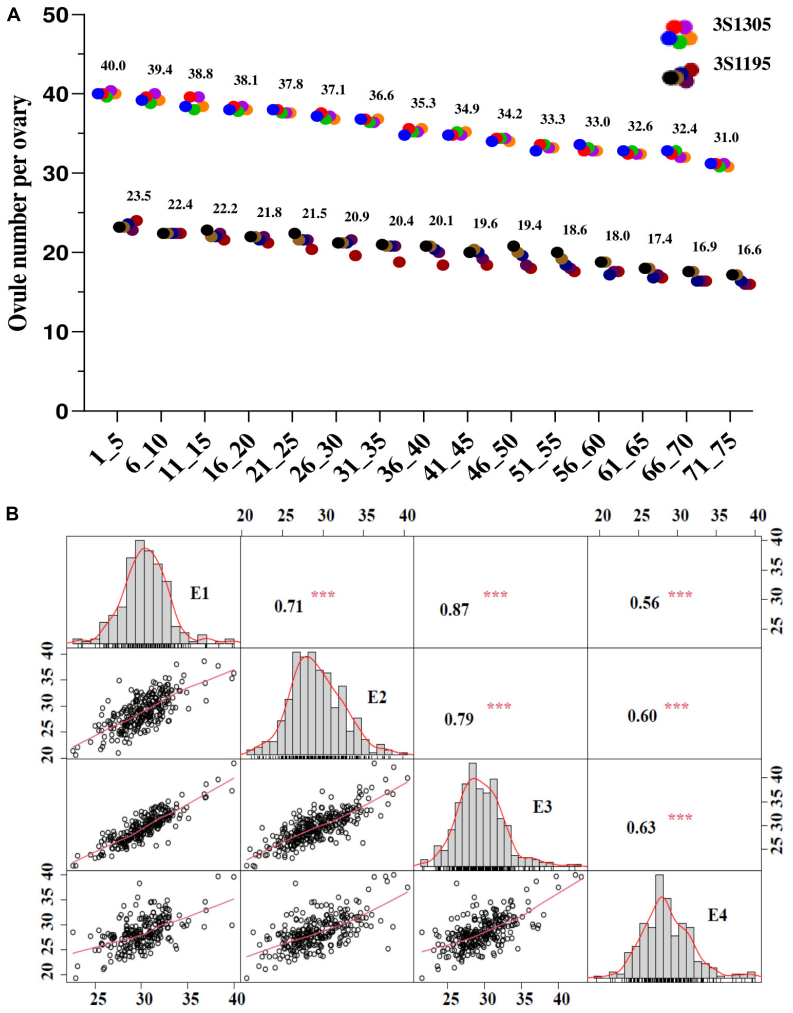


FIGURE 1
The phenotypic variation and correlation of ONPO in an association panel of oilseed rape across four investigated environments. **(A)** The ONPO in the different positions of the main inflorescence from two representative lines (3S1305 and 3S1195). The horizontal and vertical axes show the bud order and ONPO, respectively. **(B)** The frequency distribution and correlation of ONPO in four investigated environments. The diagonal line plot shows the ovule number frequency distribution. The numerals above the diagonal line are Pearson-correlation coefficient values between environments, and diagrams below the diagonal line indicate the scatter plots of the ovule number. *** Represents the significance level of $P = 0.001$.

TABLE 1 Descriptive statistic of ovule number per ovary trait in four investigated environments.

Environments	Mean ± SE	SD	Min	Max	Variance	CV (%)	Kurtosis	Skewness
20QH	30.14 ± 0.13	2.69	22.20	40.20	7.22	8.91	1.25	0.26
20YL	29.11 ± 0.12	3.54	19.70	42.00	12.52	12.16	0.22	0.34
21WC	29.64 ± 0.11	3.45	20.30	43.80	11.89	11.64	1.25	0.64
21YL	28.50 ± 0.14	3.28	19.22	40.80	10.75	11.50	1.32	0.67

20QH, 20/21YL and 21WC are the abbreviations of the 20 Qinghai, 20/21 Yangluo and 21 Wuchang environments, respectively.

The A allele's mean ONPO (30.5) was significantly greater than that (29.1) of the C allele with $P = 1.00E^{-3}$. For *qONPO.C03-2*, the two variant bases A and G of peak SNP Bn-scaff_18322_1-p806079, respectively, accounted for 82.5 and 17.5%. The A allele's mean ONPO (30.1) was significantly greater than that (29.0) of the G allele with $P = 2.80E^{-3}$.

Furthermore, relative to the leading SNPs of these association loci, a total of 1971 annotated genes were located within 500 kb or LD statistic $r^2 > 0.2$. A dozen QTL for ONPO

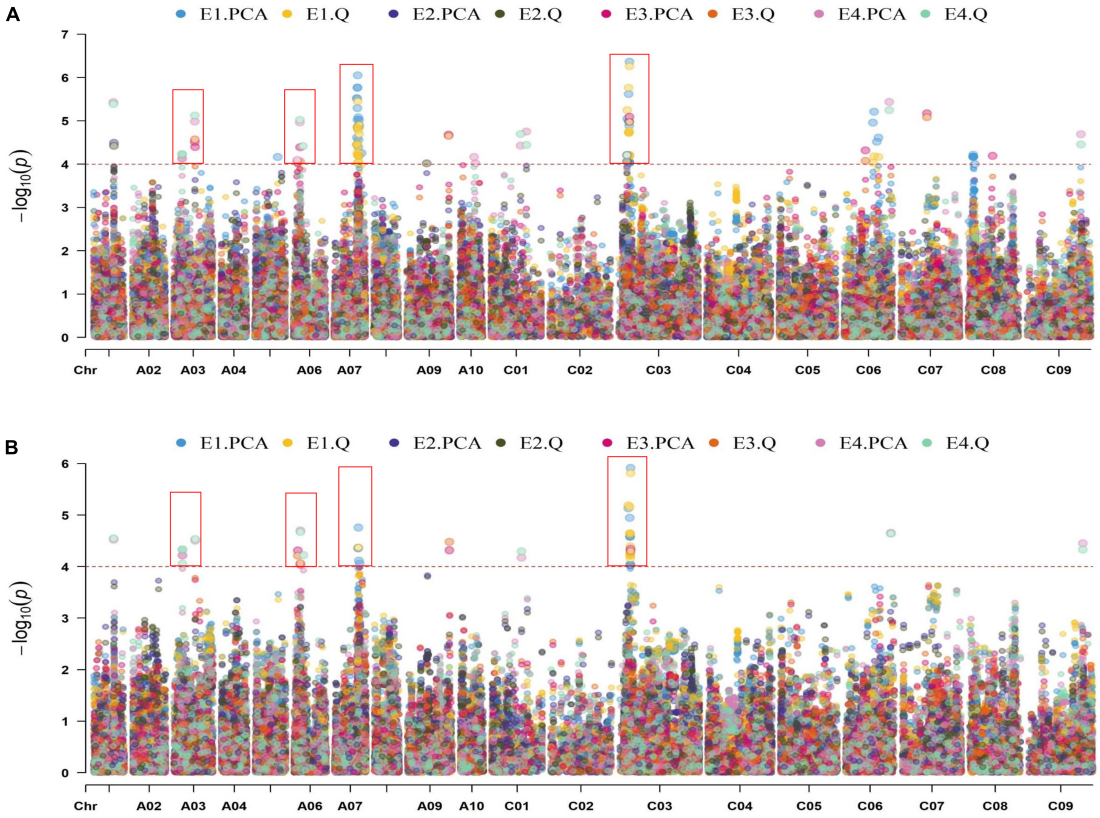


FIGURE 2
Manhattan plot of GLM **(A)** and MLM **(B)** in four environments. The horizontal and vertical axes show the length of each chromosome and $-\log_{10}P$-value, respectively. The scatters on the figure show the $-\log_{10}(P\text{-values})$ of the corresponding SNP markers. The dotted lines on the figure showed the threshold for the $-\log_{10}(P\text{-value})$. The different colors distinguished the different environments and models, as shown in the legends.

have been reported through linkage or association analysis in oilseed rape (Okada et al., 1991; Jiao et al., 2021). To determine their positional relationship with the association loci identified in the current research, comparative QTL analysis was conducted based on the reference genome of Darmor_V4.1.[3] The results indicated that all 18 association loci are not overlapped with the published QTLs of ONPO, suggesting that they are all novel loci.

Screening of lines with extreme ovule number

In order to further comparative study, a total of 26 representative lines with more or less ovule were selected based on their ONPO data in four environments. The ONPO of 13 more- and 13 less-ovule lines ranged from 32.8 to 40.1 and from 20.8 to 26.2, respectively, and the mean of the former (35.7) was significantly larger than that of the latter (24.0) with $P = 8.20\text{E}^{-13}$ (Figure 4A). As expected, the ONPO of these extreme lines displayed a significant ($P = 2.78\text{E}^{-13}$) positive correlation with SNPS, with $r = 0.617$ (Figure 4B). The difference between ONPO and SNPS of the M and L line ranged from 8.6 to 25.6 and 6.5 to 18.4, respectively (Figure 4C), and the mean of the former (18.1) was significantly larger than that of the latter (10.9). The calculated seed-setting rate of the M and L lines varied from 34.8 to 75.1% and from 29.8 to 71.1%, respectively. The average seed-setting rate of M and L lines (53.3 and 54.5%) had no significant difference, indicating that the genetic control of ONPO and seed-setting rate should be different (Figure 4D).

Comparative physiological study between extreme lines

To obtain insight into the physiological processes that affect ONPO, the leaf area, leaf photosynthetic rate, and phytohormones content were investigated and compared between the extreme lines.

3 https://www.ncbi.nlm.nih.gov/assembly/GCA_000751015.1

10.3389/fpls.2022.999790

TABLE 2 The details of 18 association loci of ovule number per ovary.

QTLs	Model	Marker	Position	−log10 (*P*)	Ad_ effect	PVE (%)	ENV
qONPO.A01-1	1,2,3,4	Bn-A01-p17498057	14,947,415	−0.734609145	0.57822	9.3	E4
qONPO.A01-2	1,2	Bn-A01-p17118647	15,506,771	−0.652511091	0.88558	6.5	E3
qONPO.A03-1	4	Bn-A03-p7234862	6,510,685	−0.636440903	0.85913	7.8	E4
qONPO.A03-2	1,2,3,4	seq-new-rs32414	16,137,692	−0.709442121	−1.0347	8.7	E4
	2	seq-new-rs32414	16,137,692	−0.659806462		6.9	E2
qONPO.A06-1	1,2,3,4	seq-new-rs23621	4,708,433	−0.671900437	0.98819	8.6	E4
	1	Bn-A06-p5616164	5,062,564	0.000040773		4.3	E2
qONPO.A06-2	2	seq-new-rs39246	7,101,109	−0.645020967	/	6.1	E4
qONPO.A07-1	1,2	Bn-A07-p15591535	17,487,704	−0.74167809	0.54459	9.3	E1
qONPO.A07-2	1,2,3,4	seq-new-rs40795	18,145,354	−0.781835256	−0.39028	10.3	E1
qONPO.A07-3	1,2	Bn-A07-p17225853	19,141,651	−0.700557723	−0.16366	8.6	E1
qONPO.A09-1	1,2,4	seq-new-rs41467	31,832,091	−0.651051124	1.15762	7.0	E2
qONPO.C01-1	1,2	seq-new-rs36223	22,560,520	−0.671818418	−0.34682	7.8	E4
qONPO.C01-2	1,2	seq-new-rs42691	27,093,843	−0.677173624	0.21882	8.1	E4
qONPO.C03-1	1,2,3,4	seq-new-rs30160	6,454,037	−0.710287855	−0.38299	9.9	E1
qONPO.C03-2	1,2,3,4	Bn-scaff_18322_1-p806079	8,231,877	−0.771941275	−0.060327	11.5	E1
	1,2,3	Bn-scaff_18322_1-p806079	8,231,877	−0.706911562		7.5	E2
qONPO.C06-1	1	seq-new-rs43825	22,653,417	−0.71681982	/	7.6	E1
qONPO.C06-2	1,2,3,4	seq-new-rs45976	34,229,155	−0.73531806	2.70296	9.3	E4
qONPO.C07-1	1,2	seq-new-rs28170	19,782,297	−0.713787513	/	7.5	E2
qONPO.C09-1	1,2,3,4	Bn-scaff_19899_1-p12334	41,095,875	−0.648738222	/	6.7	E4

E1–E4 are the codes of the above four environments 20QH, 20YL, 21WC and 21YL, respectively. 1–4 are the four models GLM-PCA, GLM-Q, MLM-PCA, and MLM-Q, respectively.

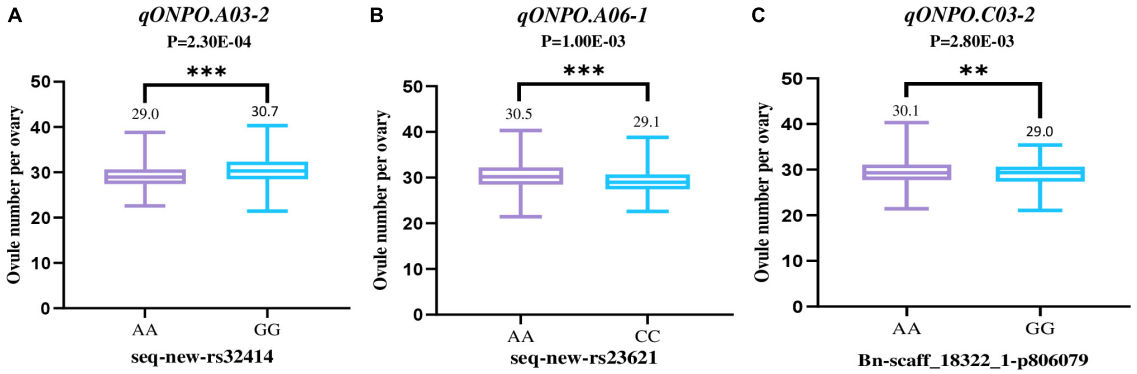


FIGURE 3
The phenotypic effects of three repeatable association loci **(A–C)**. The horizontal axis shows the haplotype of the peak SNP for the association loci. The vertical axis shows the ONPO. The violin diagram in the figure shows the mean and range of ONPO for each haplotype. The *P*-values of the *T*-test between the ONPO of corresponding haplotypes were also shown at the top of figure. ** and *** represent the significance level of $P = 0.01$ and 0.001, respectively.

The photosynthetic rates of 13 more- and 13 less-ovule lines varied from 26.36 to 38.82 mol/m^2/s and from 27.47 to 38.59 mol/m^2/s, respectively (Figure 5A). The mean of the former (32.88 mol/m^2/s) had no significant difference from that of the latter (33.97 mol/m^2/s). As expected, the correlation between leaf photosynthetic rate and ONPO of these lines was very low and not significant ($r^2 = 0.0002$).

The leaf area of the 13 more-ovule and 13 less-ovule lines ranged from 152.32 to 306.12 cm^2 and from 144.03 to 247.34 cm^2, respectively (Figure 5B). It should be noted that the average leaf area of 13 more-ovule lines (210.40 cm^2) was not significantly larger than that (199.97 cm^2) of the 13 less-ovule lines ($P = 5.13E^{-01}$). The correlation between leaf area and ONPO of these lines were also low and not significant

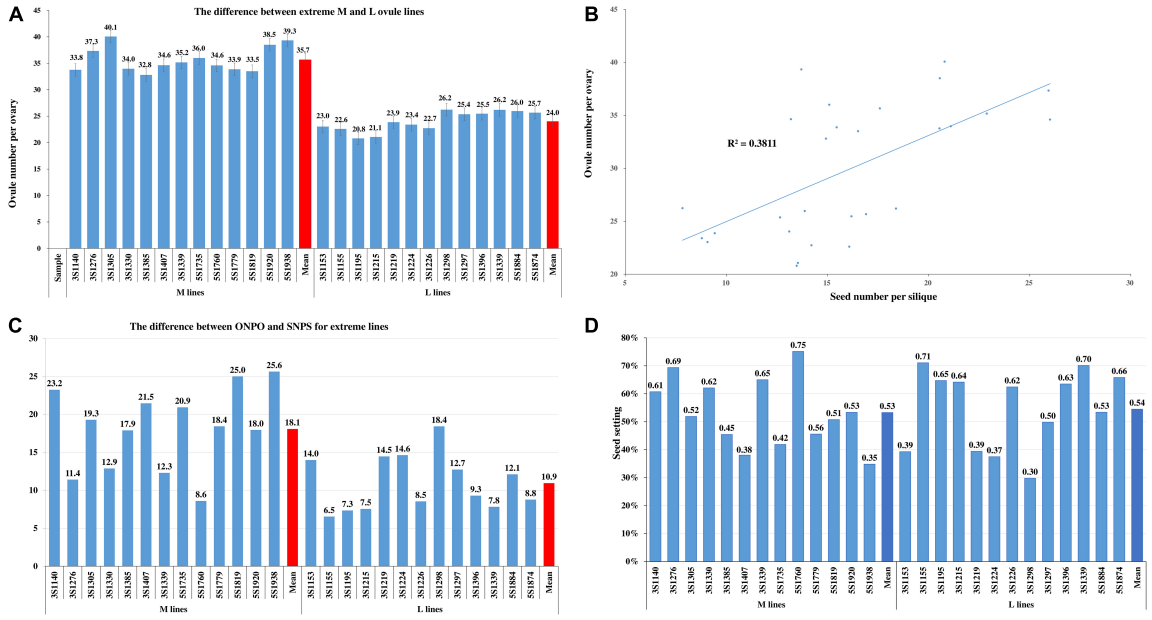


FIGURE 4
The ONPO of extreme lines and its relationship with SNPS. **(A)** The of ONPO between 13 more- and 13 less-ovule lines. The horizontal and vertical axes show the line number and ovule number per ovary, respectively. The columns of different heights indicate the number of ovules. The bar on each column represents the standard deviation. **(B)** The correlation between ONPO and SNPS of 26 extreme lines. The horizontal and vertical axes show SNPS and ONPO, respectively. The positions of 26 dots show the corresponding ovule and seed number of them. The 26 pairs of data were used to fit the trend line with R^2. **(C)** The difference between ONPO and SNPS of extreme lines. **(D)** The seed-setting rate of 26 extreme lines.

(r^2 = 0.0054), suggesting that leaf area may not affect ONPO (Supplementary Figure 2).

The contents of BA, GA4, and JA in M lines (255.5, 262.5, and 28.6 ng/g) were significantly higher than those in L lines (181.5, 186.9, and 15.8 ng/g). In contrast, L lines exhibited significantly higher contents of ABA, IAA and SA (49.0, 174.9, and 151.7 ng/g) than M lines (35.2, 117.2, and 133.5 ng/g). There was no significant difference in the contents of IBA and SA between the two types of extreme lines (Figure 5C). These data suggested that the content difference of the above phytohormones should be associated with ONPO difference between M and L lines (Supplementary Table 4).

Comparative transcriptomic analysis between extreme lines

To investigate how the ovule number difference formed at the molecular level, a comparative transcriptome analysis was performed using the ovaries of M and L extreme lines at the ovule initiation stage. The summary of the transcriptomic data produced by the Illumina sequencing platform is presented in Supplementary Table 5. After filtering out low-quality reads, 143,282,598 and 148,303,824 total reads and 143.28 M to 148.3 M clean reads were obtained from RNA sequencing of three repeats of more- and less-ovule lines, respectively (Supplementary Table 5). The mapped reads rate of the L and M groups varied from 85.5 to 90.9% and from 89.5 to 92.8%, with a mean of 89.0 and 92.0%, respectively. The Q30 percentage of the L and M group ranged from 89.1 to 90.4% and from 89.6 to 90.1%, with a mean of defined 89.9 and 89.8%, respectively. The L and M group's guanine and cytosine (GC) contents varied from 45.7 to 45.8% and from 45.4 to 46.3%, with a mean of 45.8 and 45.9%, respectively.

The gene expression density of six samples was similar (Supplementary Figure 3A). As expected, the correlation coefficients among the three repeats of the same group (M or L group) are larger than those between the two groups (M and L) (Supplementary Figure 3B). The FPKM value of each expressed gene was also calculated, and its distribution was similar among the six samples (Supplementary Figure 3C). We carried out the hierarchical clustering analysis (HCA) and principal component analysis (PCA) to differentiate between the two groups. The results showed that the transcriptome of M and L lines was very different (Figures 6A,B). We constructed the volcano plot to determine the significantly expressed genes for the identification of DEGs (Figure 6C). A total of 4,449 DEGs were identified from the 95,791 expressed genes in two groups, containing 2,095 up-regulated and 2,354 down-regulated genes between two types of lines (Figure 6D). To validate the accuracy/reliability of RNA-seq analysis, ten representative DEGs were randomly selected to validate through qRT-PCR analyses. Except for *BnaC07g41610D*, most of the selected genes displayed similar expression patterns with the

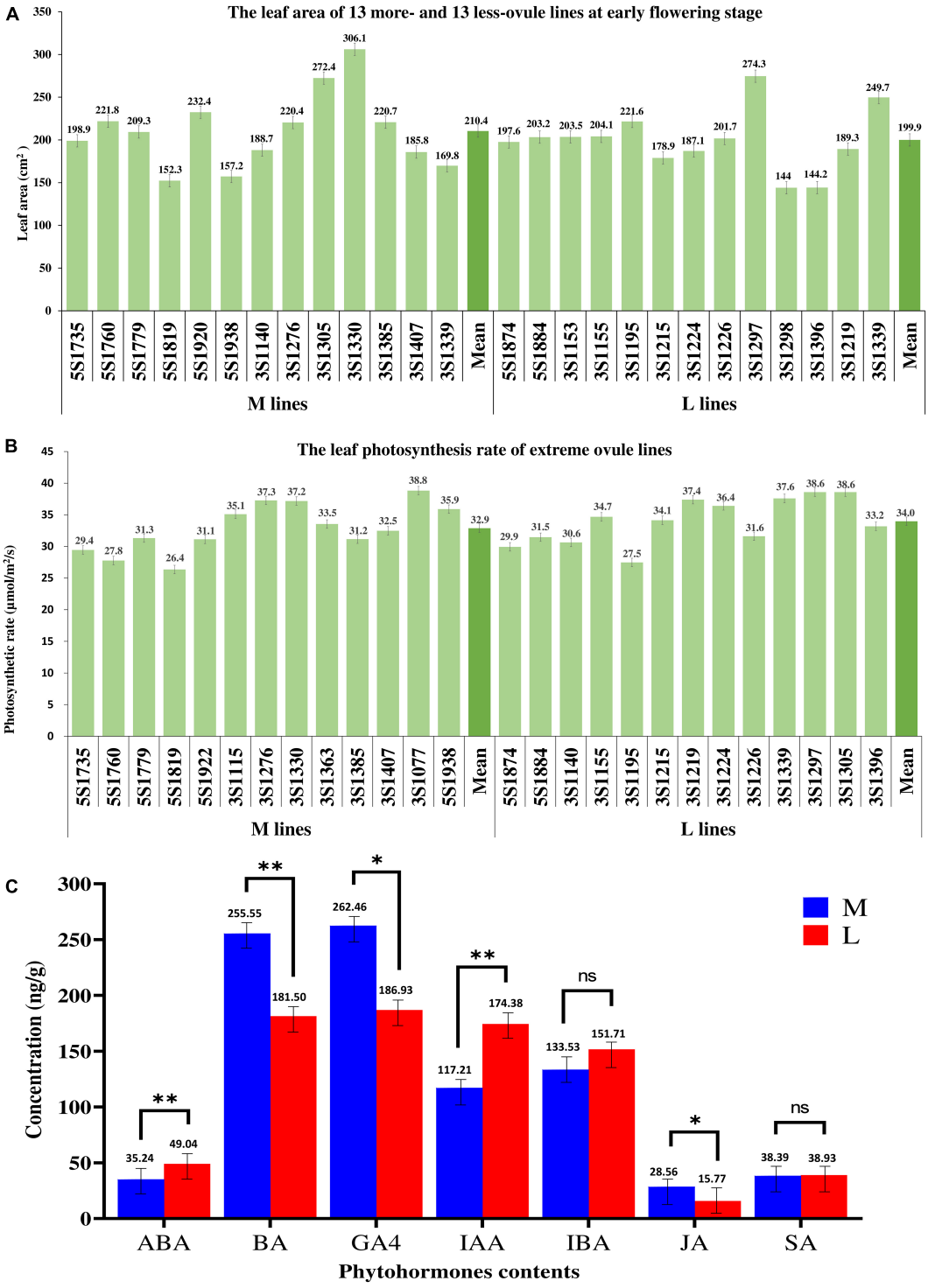


FIGURE 5
The comparison of three physiological indexes between more- and less-ovule lines. **(A,B)** The leaf area leaf photosynthetic rate of 13 more- and 13 less-ovule lines. The horizontal axis shows the line number, and the vertical axis shows **(A)** leaf area or **(B)** leaf photosynthetic rate. The columns of different heights show the leaf area and photosynthetic leaf rate of different lines. The numerals and bars on each column represent the mean and standard deviation, respectively. **(C)** Comparison of phytohormones content between more- and less-ovule lines. The horizontal and vertical axes show the phytohormone types and their content, respectively. The columns of different heights show the contents of different lines. The numerals and bars on each column represent the mean and standard deviation, respectively. * and ** represent the significance level of $P = 0.05$ and 0.01, respectively.

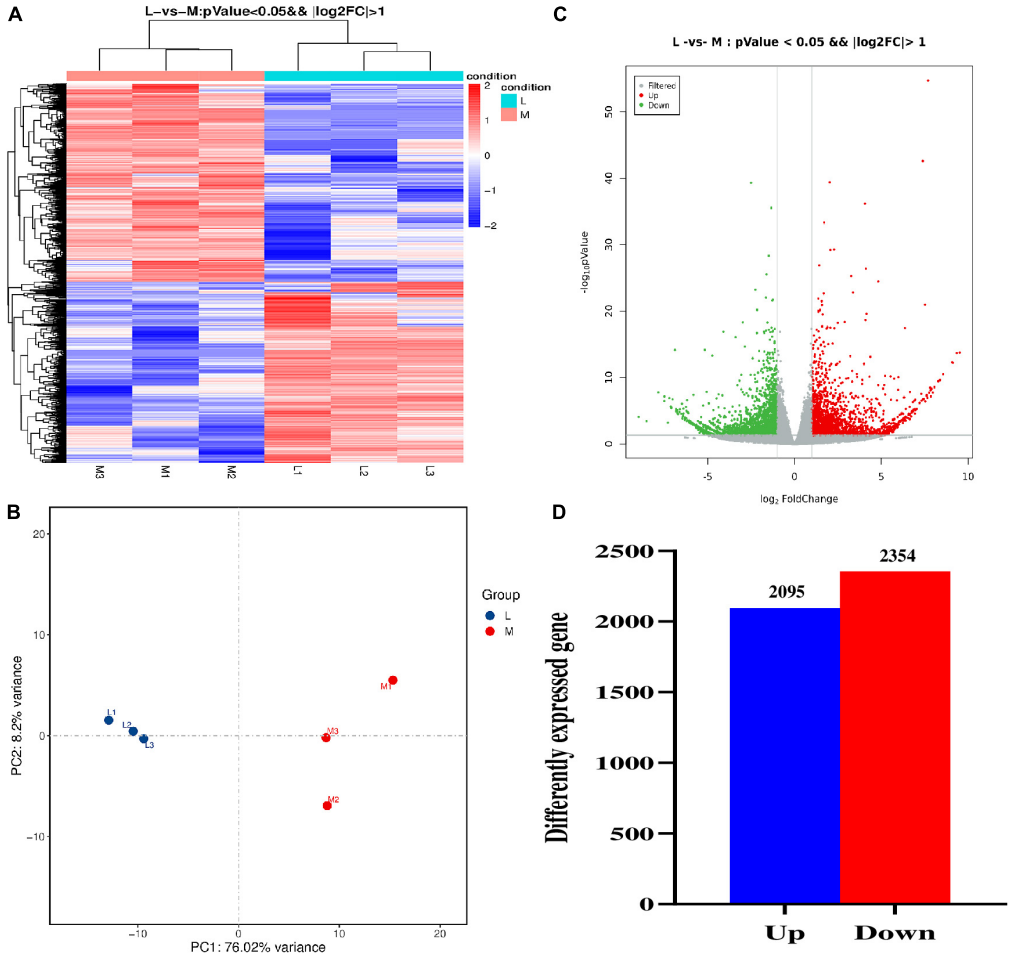


FIGURE 6
The comparative transcriptomic analysis between more- and less-ovule lines. **(A)** The heat map of DEGs between more- and less-ovule lines. The color column indicates the normalized signal value. The more- and less-ovule lines were distinguished using the different colors. L1 to L3 and M1 to M3 represent the three repeats and less and more-ovule lines. **(B)** Principal component analysis of gene expression level of the three repeats of more- and less-ovule lines. The horizontal and vertical axes show the variance of PC1 and PC2, respectively. The more- and less-ovule lines were distinguished using the different colors. **(C)** The volcano map of expressed genes in more- and less-ovule lines. The horizontal and vertical axes show $\log_2$ Fold Change and $-\log_{10}$(*P*-value), respectively. **(D)** The statistic of DEGs number between more- and less-ovule lines. The horizontal and vertical axes show the up/down pattern and number of DEGs, respectively.

transcriptomic data, although the fold of change varied between the two methods (**Supplementary Figure 4**).

The functional categorization results of the DEGs were further confirmed *via* gene ontology (GO) and KEGG analysis (**Figure 7**). A total of 4,449 DEGs were subjected to an enrichment analysis for GO annotation terms. The biological process included adhesion, regulation, biogenesis, cellular and developmental process, reproductive process and rhythmic process, etc. These DEGs were enriched in cell junction, extracellular matrix, nucleotide, organelle, and virion for cellular components. For molecular function, these DEGs were enriched in Antioxidant activity, catalytic activity, molecular structure-activity, transporter, and receptor activity, etc. Moreover, we performed a KEGG enrichment analysis after identifying DEGs in two pools to characterize the DEGs. The results showed that these DEGs were enriched in 19 biological pathways, including environmental adaptation, nucleotide metabolism, carbohydrate metabolism, translation, genetics, biosynthesis of other secondary metabolism, and proteins, etc.

Characterization of differentially expressed genes

To reveal the molecular mechanisms that involve the DEGs, they were submitted to the super viewer[4], which displayed the total number and frequency as well as the *P*-value for each class

4 http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer.cgi

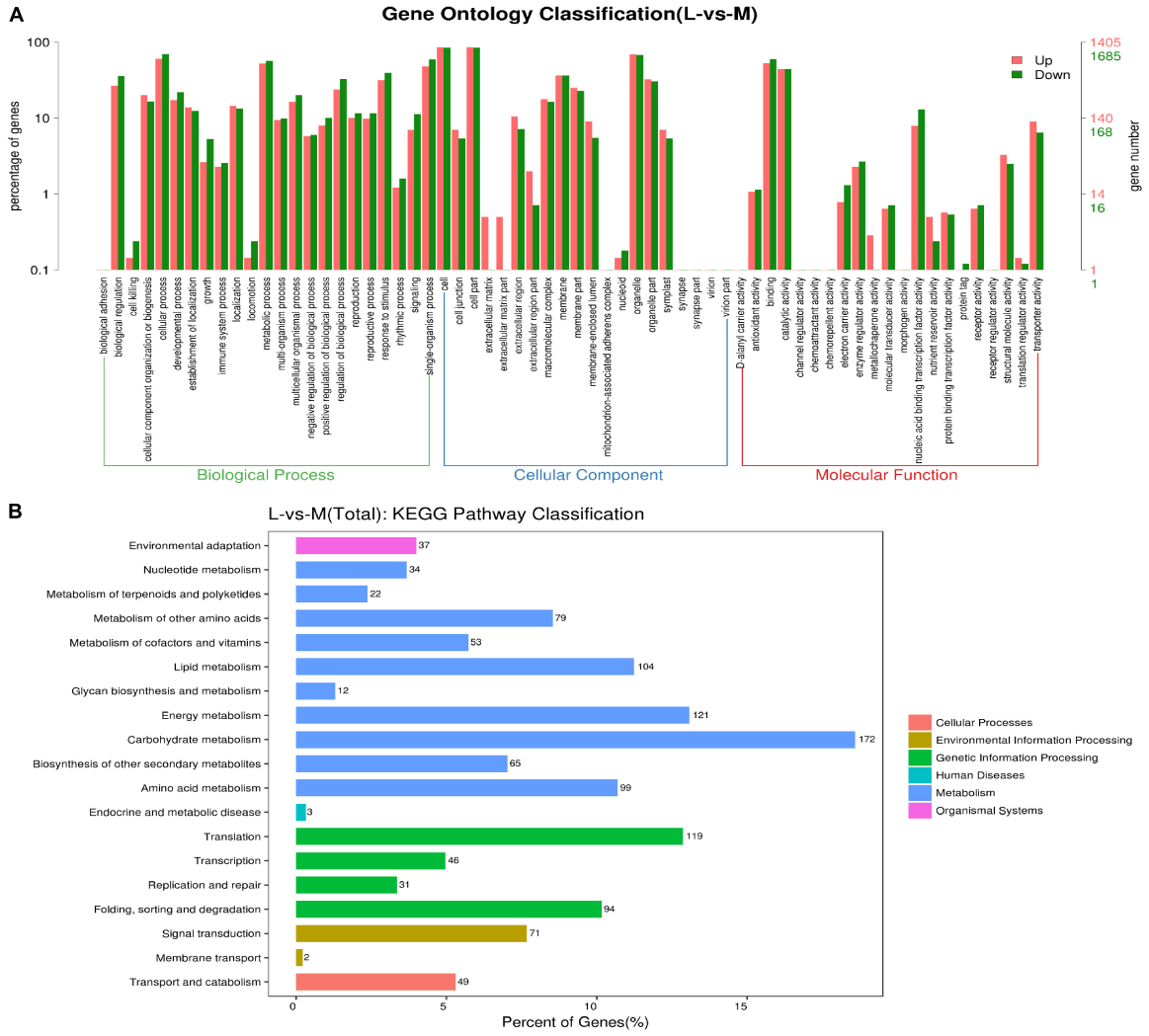


FIGURE 7
The functional categorization of differentially expressed genes. **(A)** GO enrichment histogram of DEGs. The horizontal axis shows three main categories (biological process, cellular component, and molecular function), each containing several subclasses. The vertical axis shows the number and proportion of DEGs in up and down-regulation. **(B)** KEGG enrichment histogram of DEGs. The horizontal and vertical axes show the number/percent and type of DEGs, respectively.

(Supplementary Table 6). The results indicated that the DEGs are involved in diverse metabolic processes, including DNA (57), RNA (540), photosynthesis (PS, 79), transport (176), protein (489), signaling (196), stress (121), cell (114), cell wall (97), hormone metabolism (74), etc.

Of the 540 DEGs in the largest group of RNA classes, 341 (63.1%) were involved in transcription factors, indicating the importance of TF in regulating ovule number. These transcription factors DEGs were from 47 types, including unclassified TFs (35), MYB (33), bHLH (33), Putative transcription regulator (21), Homeobox (21), C2H2 zinc finger (20), AP2/EREBP (15), WRKY (11), bZIP (11), B3 TFs (10), C2C2 (Zn) DOF zinc finger (9), C2C2 (Zn) GATA (9), G2-like TFs (9), MADS box (8), ARR (8), Trihelix (6), Aux/IAA (6), Chromatin Remodeling Factors (5), General Transcription (5), SET-domain (4), C2C2 (Zn) CO-like (4), AS2 (4), Zf-HD (3), YABBY (3), HSF (3), CCAAT box binding factor (3), E2F/DP TFs (3), DNA methyl transferases (3), FHA TFs (2), Global TFs (2), Histone acetyltransferases (2), sigma like plant (2), RNA regulation of transcription (2), ABI3/VP1 family (2), C2C2 (Zn) Alfin-like (2), TCP (2), Methyl binding domain proteins (2), Nucleosome/chromatin (2), Polycomb Group (2), Zn-finger (CCHC) (2), GRP (1), GRF zinc finger (1), AT-rich interaction domain (1), AtSR (1), JUMONJI (1), NAC (1), NIN-like bZIP (1), PWWP domain protein (1), Silencing Group (1), SNF7 (1) and Pseudo ARR (1).

A total of 489 DEGs were involved in the second largest protein class group, including protein degradation

ubiquitin (123), protein posttranslational modification kinase (73), synthesis. ribosomal protein (56), protein targeting unspecified (15), protein degradation (14), protein degradation cysteine protease (14), synthesis ribosomal biogenesis (14), protein synthesis initiation (14), protein degradation serine protease (11), protein degradation AAA type (8), protein synthesis elongation (7), protein degradation subtilizes (7), protein targeting nucleus (6), protein folding (6), protein targeting mitochondria (6), protein targeting chloroplast (4), protein targeting golgi pathway (4), protein targeting vacuole (4), protein targeting plasma membrane (2), protein targeting peroxisomes (2), protein targeting unknown (2), protein degradation autophagy (2), protein degradation metalloprotease (2), tyrosine-tRNA ligase (1), methionine-tRNA ligase (1), aspartate-tRNA ligase (1), cysteine-tRNA ligase (1), arginine-tRNA ligase (1), asparagine-tRNA ligase (1), bifunctional aminoacyl-tRNA synthetize (1), lysine-tRNA ligase (1), alanine-tRNA ligase (1), valine-tRNA ligase (1), activation (1), protein targeting ER pathway (1) and protein glycosylation (1).

It should be noted that a set of 74 DEGs were found to be associated with metabolic and signaling pathways of multiple hormones, including AUX (27), ETH (11), JA, (9), SA (9), BR (6), CK (5), ABA (4), and GA (3), showing the complex role and interactions of phytohormones (**Figure 8**). Of the 74 DEGs, many were associated with the synthesis or degradation of phytohormones, potentially contributing to the hormone concentration difference between more- and less-ovule lines (**Supplementary Table 7**).

To establish a link between the DEGs and ovule number, they were BLAST against the reported ovule number genes, mainly from *Arabidopsis* (Qadir et al., 2021). Of these DEGs 12 are homologous to known ovule number genes such as *CUC1* (*BnaA01g28990D*), *REV* (*BnaA02g06170D*), *CRF3* (*BnaA03g12320D*), *CRF6* (*BnaC06g42850D*), *CYP85A2* (*BnaC02g38080D*), *GA20OX* (*BnaC03g75160D*), *AP2* (*BnaCnng39690D*), *PAN* (*BnaC06g30310D*), *SHP1* (*BnaC08g29520D*), *CRC* (*BnaA07g27740D*), *INO* (*BnaC07g13110D*), and *HEMN1* (*BnaCnng15790D*). More than half of DEGs are involved in the know pathways of ovule number regulation, including development, phytohormones, transcription factors, protein, signaling (phosphorelay, sugar and nutrient physiology, G-protein and receptor kinase), and micro-RNA. These results highly suggested that the identified DEGs were highly associated with ovule number (Qadir et al., 2021).

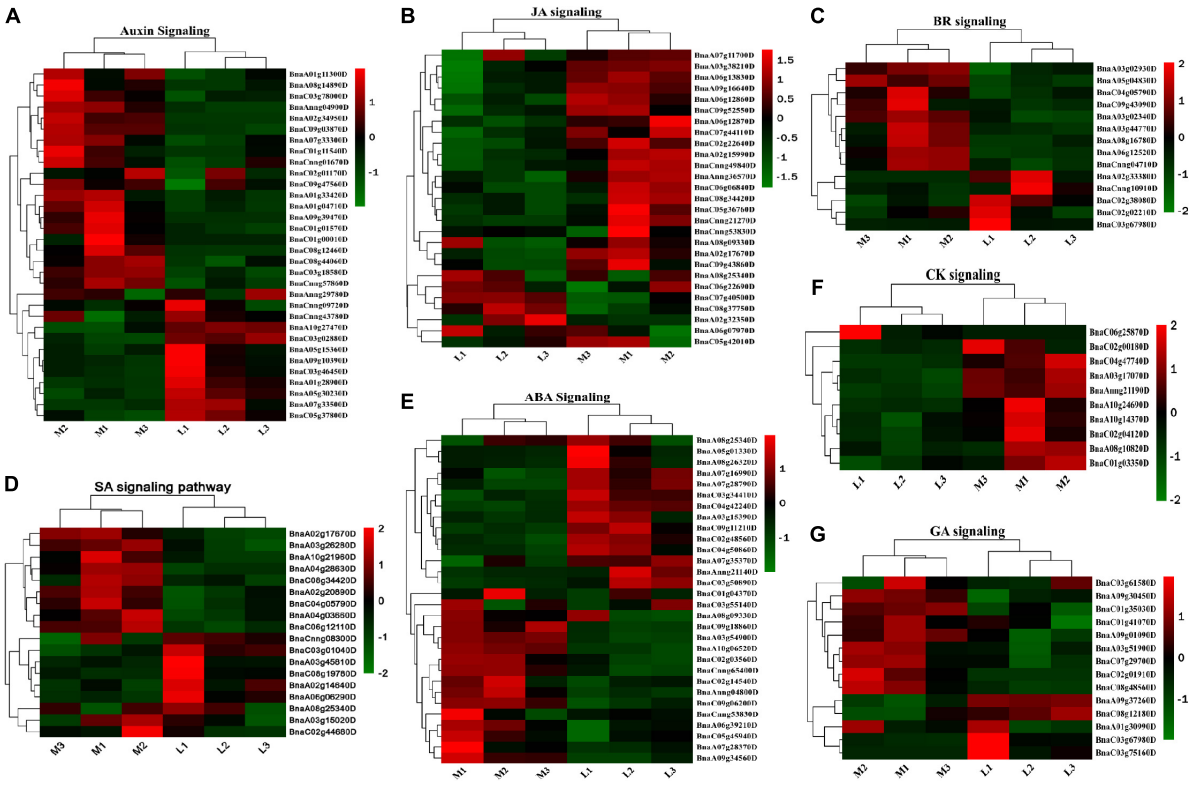


FIGURE 8
The demonstration of differentially expressed genes (DEGs) involved in the phytohormones pathway. **(A–G)** Heatmap of DEGs in ABA, JA, CK, SA, BR, IAA, and GA pathway. The sample and DEG no were shown on the bottom right of the figure, respectively. The color column in the legend represents the gene expression level.

Identification of candidate genes through the integration of genome-wide association study and RNA-seq

To identify the candidate genes underlying the natural variation of ONPO in oilseed rape, a Venn diagram was constructed between 1,971 annotated genes within the genomic region of association loci and the 4,449 DEGs, which resulted in 73 overlapping genes (**Supplementary Figure 5**). Integrating functional annotation information, six genes underlying association loci on A03, A06, C01, and C06 chromosomes were selected, including *BnaA03g14600D*, *BnaA03g33420D*, *BnaA06g08920D*, *BnaA06g13210D*, *BnaC01g25840D*, and *BnaC03g16210D*.

Discussion

In angiosperms, ovules are critical organs as they represent the direct progenitors of seeds (Endress, 2011; Qadir et al., 2021). Therefore, dissecting the genetic basis of ovule numbers will provide valuable information for the targeted improvement of seed crops' yield (Banks et al., 2010; Qadir et al., 2021). In the present study, the GWAS, transcriptomic and physiological analyses were conducted to identify causal loci and candidate genes underlying ovule number variation in *B. napus*.

Great potential for genetic improvement of ovule and seed number

The ONPO of the association population displayed a large variation from 19.2 to 43.8 (mean≈29) across the four investigated environments. This range (>20) was obviously larger than the previous reports (<13) in *B. napus* (Ali, 2018; Chen et al., 2019; Khan et al., 2019), which represented a valuable resource for the genetic improvement of this trait. It should be noted that the reported ONPO of oilseed cultivars (Li et al., 2014; Yang et al., 2017) was close to the mean of the association population, suggesting that this trait has hardly been selected during oilseed rape breeding. The large difference (>10) between the ONPO of current cultivars and the max of germplasm indicated the great potential for the genetic improvement of ONPO in oilseed rape.

It should be noted that the variance associated with genotype (8.073) was 21.3 times greater than that of the environment (0.379), suggesting that genotype rather than the environment is the primarily determinant of ovule number variation in oilseed rape. As expected, the calculated broad-sense heritability of ONPO in the current study (91.9%) was higher than those of most yield components and related traits reported in *B. napus* (Cai et al., 2014; Wang et al., 2020; Hussain et al., 2021), which suggested that it was mainly governed by genotype.

In addition, the seed-setting rates of 26 extreme lines were accurately calculated using ONPO and SNPS measured in the same experiment, which varied greatly from 29.8 to 75.1%, with a mean of 54.0%. Similarly, the seed-setting rate of 36 DH lines derived from C010 × C001 varied from 38.3 to 82.2%, with a mean of 56.1% (Chen et al., 2019). It should be noted that the reported seed-setting rates of oilseed rape cultivars, including ZY-50 (69.3%) and Zhongshuang11 (75.4%) were all larger than the corresponding mean of 26 extreme lines (Yang et al., 2017; Ali, 2018), which suggested that it should be strongly selected in breeding. The large gap between SNPS and ONPO highly indicated the great potential for the genetic improvement of seed number (Li et al., 2017).

Novel loci associated with ovule number per ovary variation

Although a dozen linkage and/or association mapping studies have been conducted for seed number per silique in oilseed rape (Shi et al., 2015; Lu et al., 2017; Yang et al., 2017), rarely was about ovule number per ovary. Eight association loci were identified in an association panel of 521 oilseed rape accessions genotyped with the *Brassica* 60 K SNP array, which was distributed on A03, A09, A10, C02, C04, and C05 chromosomes, explaining 1.22–6.40% of phenotypic variance (Khan et al., 2019). Five QTLs were detected using a DH population of 180 lines derived from inbred lines 7–5 and ZY50, which were distributed on A03, A07, A10 and C06 chromosomes, explaining 1.9–17.38% of the phenotypic variance (Ali, 2018).

In the current study, a total of 18 association loci were identified using an association panel of 327 lines genotyped with *Brassica* 50 K SNP array. It should be noted that the physical positions of all these 18 association loci were different from the reported QTL of ovule number, suggesting all of them to be novel. More importantly, three association loci were repeatedly identified in the different environments, suggesting it to be the important target for maker-assisted selection (since breeder can't see ovule number by eye) and further gene cloning.

Physiological basis of ovule number variation

The previous studies showed that the ONPO was affected by flower size (Wetzstein et al., 2013; Cucinotta et al., 2020), flower position, nutrient availability (Chalcoff and Aizen, 2016), and phytohormones (Gomez et al., 2018; Barro-Trastoy et al., 2020). Therefore, several related physiological indexes

(including flower order, leaf area and photosynthetic rate, and phytohormones content) were measured and compared between the M and L ovule lines.

Obviously, the ONPO of two representative extreme lines decreased with the bud positions from the bottom to top of the main inflorescence, which was consistent with a previous finding that ovule number decreased from the basal to distal of racemes in *Hosta ventricosa* (Cao et al., 2011). This result could be explained by the distance between the top flowers and the source organs of assimilates, which is highly accordant with the resource competition hypothesis. Whereas both leaf area and photosynthetic rates of more- and less-ovule lines have no significant difference between the more- and less-ovule lines, further study should focus on the leaf area index because it's more representative of the resource. It has been well known that plant hormones play an important regulatory role in regulating ovule initiation and number (Barro-Trastoy et al., 2020; Cucinotta et al., 2020; Qadir et al., 2021; Yang and Tucker, 2021). In the present study, the contents of several types of endogenous phytohormones (ABA, BA, GA4, IAA, JA) have significant differences between more- and less-ovule lines. Highly accordant with this, the transcriptomic analysis identified 74 DEGs that are involved in the phytohormones pathway. More importantly, many DEGs are involved in the metabolic (such as synthesis and degradation) process of the same type of phytohormones (ABA, BR, CK, ETH, GA, IAA, JA, SA). In the future, it's worth investigating the response of ovule number on different types of phytohormones using the representative lines in the oilseed rape crop.

Association of identified deferentially expressed genes with ovule number

The 4,449 DEGs were enriched into a dozens of classes, including RNA (540), protein (489), signaling (196), transport (176), stress (121), cell (114), cell wall (97), photosynthesis (79), hormone metabolism (74), DNA (57), etc. This confirmed the complexity of the ovule initiation and development process at the transcriptome level. Of the 540 DEGs enriched in the largest class of RNA, most belonged to the transcription factors, which was consistent with the previous finding that Transcription factors (TFs) have been reported to play crucial roles in the reproductive development of flowering plants (Yang and Tucker, 2021). In addition, a total of 74 DEGs were involved in the metabolic and signaling pathway of several types of phytohormones, which was accordant with the different content of several types of phytohormones. Together, these results support the importance of phytohormones in ovule initiation and development (Barro-Trastoy et al., 2020; Qadir et al., 2021; Yang and Tucker, 2021; Yu et al., 2022). More importantly, about half of the DEGs belonged to the same functional category of the identified ovule number regulating genes. Especially, 12 DEGs are homologous to the known ovule number genes. These results highly suggested that the identified DEGs are associated with ovule number. Further studies are worthy of validating the function of the 12 homologous ovule number DEGs in oilseed rape by over-expression, RNAi, or gene editing.

Candidate genes underlying ovule number per ovary variation

Integrating genome-wide association and transcriptome analysis is an efficient strategy for discovering candidate genes of complex traits (Li et al., 2020). In the current study, integrating association loci, DEGs and functional annotation, six candidate genes were identified, including *BnaA03g14600D, BnaA03g33420D, BnaA06g08920D, BnaA06g13210D, BnaC01g25840D,* and *BnaC03g16210D*. *BnaA03g33160D* is homologous to *Arabidopsis EIF4A1* that encodes RNA helicase, whose mutant decreased in both ovule number and fertility (Bush et al., 2015). *BnaA01g23050D* is homologous to *Arabidopsis HAP13*, which encodes μ1 adaptin component of the heterotetrameric protein complex that regulates protein sorting at the *trans*-Golgi network/early endosome. Its mutant displayed defects in outer integument growth as well as reduced ovule number (Wang et al., 2016). *BnaA03g33770D* is homologous to *Arabidopsis* transcription factor *CUC1*, which functions redundantly with *CUC2* and *CUC3* to regulate ovule initiation. The double mutant of *CUC1* and *CUC2* showed reduced ovule number (Cucinotta et al., 2018). These important candidate genes will be subjected to comparative sequencing and further functional validation in oilseed rape.

Conclusion

In summary, this study investigated the genetic, physiological, and transcriptomic basis for the natural variation of ONPO in oilseed rape using an association population and extreme lines with more and less ovules. A wide variation in ovule number exists in oilseed rape, and this variation is primarily attributed to the genotype. Through the combination of genome-wide association and transcriptomic analysis, a total of 18 novel association loci and six candidate genes were identified, which provide a solid basis for maker-assisted selection and further gene cloning.

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://www.ncbi.nlm.nih.gov/, PRJNA820145.

Author contributions

HW: formal analysis, funding acquisition, and supervision. JS: conceive and design the experiments, project administration, and revise the manuscript. MQ: perform the experiment, analyze the data, write the manuscript, and visualization. LQ: investigate the leaf area and photosynthetic rate. JY and NA: help to analyze the data. XW: resources and revise the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Key Research Program and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ZDRW202105), the Major Project of Hubei Hongshan Laboratory (2021HSZD004), Central Public-interest Scientific Institution Basal Research Fund (Y2022CG08), the Agriculture Research System of MOF and MARA of China (CARS-13), and the Agricultural Science and Technology Innovation Project (CAAS-ASTIP-2013-OCRI).

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.2022.999790/full#supplementary-material

References

Ahmad, M., Waraich, E. A., Skalicky, M., Hussain, S., Zulfiqar, U., and Anjum, M. Z. (2021). Adaptation strategies to improve the resistance of oilseed crops to heat stress under a changing climate: An overview. *Front. Plant Sci.* 12:767150. doi: 10.3389/fpls.2021.767150

Ali, A. (2018). *Quantitative trait loci (QTL) studies for the number of ovules and seeds per pod in Brassica napus* L. Dissertation/master's thesis. Wuhan, HB: Huazong Agricultural University.

Banks, H., Himanen, I., and Lewis, G. P. (2010). Evolution of pollen, stigmas and ovule numbers at the caesalpinioid-mimosoid interface (*Fabaceae*). *Bot. J. Linn. Soc.* 162, 594–615. doi: 10.1111/j.1095-8339.2010.01038.x

Barro-Trastoy, D., Carrera, E., Baños, J., Palau-Rodríguez, J., Ruiz-Rivero, O., Tornero, P., et al. (2020). Regulation of ovule initiation by gibberellins and brassinosteroids in tomato and *Arabidopsis*: Two plant species, two molecular mechanisms. *Plant J.* 102, 1026–1041. doi: 10.1111/tpj.14684

Bencivenga, S., Simonini, S., Benková, E., and Colombo, L. (2012). The transcription factors *BEL1* and *SPL* are required for cytokinin and auxin signaling during ovule development in *Arabidopsis*. *Plant Cell* 99, 2886–2897. doi: 10.1105/tpc.112.100164

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., et al. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* 115, 591–602. doi: 10.1016/S0092-8674(03)00924-3

Bilgrami, S., Liu, L., Farokhzadeh, S., Najafabadi, A. S., Ramandi, H. D., Nasiri, N., et al. (2022). Meta-analysis of QTLs controlling seed quality traits based on QTL alignment in *Brassica napus*. *Ind. Crops Prod.* 176:114307. doi: 10.1016/j.indcrop.2021.114307

Burd, M., Ashman, T. L., Campbell, D. R., Dudash, M. R., Johnston, M. O., Knight, T. M., et al. (2009). Ovule number per flower in a world of unpredictable pollination. *Am. J. Bot.* 96, 1159–1167. doi: 10.3732/ajb.0800183

Bush, M. S., Crowe, N., Zheng, T., and Doonan, J. H. (2015). The RNA helicase, *eIF4A-1*, is required for ovule development and cell size homeostasis in *Arabidopsis*. *Plant J.* 84, 989–1004. doi: 10.1111/tpj.13062

Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., et al. (2014). Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 127, 85–96. doi: 10.1007/s00122-013-2203-9

Cao, G., Xue, L., Li, Y., and Pan, K. (2011). The relative importance of architecture and resource competition in allocation to pollen and ovule number within inflorescences of *Hosta ventricosa* varies with the resource pools. *Ann. Bot.* 107, 1413–1419. doi: 10.1093/aob/mcr085

Chalcoff, V. R., and Aizen, M. A. (2016). Pollination unpredictability and ovule number in a South-Andean Proteaceae along a rainfall gradient. *Aust. J. Bot.* 64, 8–14. doi: 10.1071/BT15016

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

Chen, W., Zu, F., Luo, Y. Q., Zhao, K. Q., Zhang, J. K., Zhang, G. J., et al. (2019). Factors affecting the seed number of single silique in *Brassica napus*. *Chin. J. Oil Crops Sci.* 41, 331–339. doi: 10.7505/j.issn.1007-9084.2019.03.004

Clarke, J. M., and Simpson, G. M. (1978). Influence of irrigation and seeding rates on yield and yield components of *Brassica napus*. *J. Plant Sci.* 58, 731–737.

Cucinotta, M., Di Marzo, M., Guazzotti, A., de Folter, S., Kater, M. M., and Colombo, L. (2020). Gynoecium size and ovule number are interconnected traits that impact seed yield. *J. Exp. Bot.* 71, 2479–2489. doi: 10.1093/jxb/eraa050

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

Endress, P. K. (2011). Angiosperm ovules: Diversity, development, evolution. *Ann. Bot.* 107, 1468–1489. doi: 10.1093/aob/mcr120

Galbiati, F., Sinha Roy, D., Simonini, S., Cucinotta, M., Ceccato, L., Cuesta, C., et al. (2013). An integrative model of the control of ovule primordia formation. *Plant J.* 76, 446–456. doi: 10.1111/tpj.12309

Gomez, M. D., Barro-Trastoy, D., Escoms, E., Saura-Sańchez, M., Sańchez, I., Briones-Moreno, A., et al. (2018). Gibberellins negatively modulate ovule number in plants. *Development* 145:dev163865. doi: 10.1242/dev.163865

Helal, M. M. U., Gill, R. A., Tang, M., Yang, L., Hu, M., Yang, L., et al. (2021). SNP-and haplotype-based GWAS of flowering-related traits in *Brassica napus*. *Plants* 10:2475. doi: 10.3390/plants10112475

Hu, M., Li, X., Wang, Z., You, Q., and J, L. (2017). Effects of sowing date on biomass and nutrient accumulation of oilseed rape as green manure. *Hubei Agri. Sci.* 55, 657–660.

Hu, Q., Hua, W., Yin, Y., Zhang, X., Liu, L., Shi, J., et al. (2017). Rapeseed research and production in China. *Crop J.* 5, 127–135. doi: 10.1016/j.cj.2016.06.005

Hussain, Q., Zhan, J., Liang, H., Wang, X., Liu, G., Shi, J., et al. (2021). Key genes and mechanisms underlying natural variation of silique length in oilseed rape (*Brassica napus* L.) germplasm. *Crop J.* 8, 617–626. doi: 10.1016/j.cj.2021.08.010

Jiao, Y., Zhang, K., Cai, G., Yu, K., Amoo, O., Han, S., et al. (2021). Fine mapping and candidate gene analysis of a major locus controlling ovule abortion and seed number per silique in *Brassica napus* L. *Theor. Appl. Genet.* 134, 2517–2530. doi: 10.1007/s00122-021-03839-6

Khan, S. U., Yangmiao, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., et al. (2019). Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in *Brassica napus* L. *Ind. Crops Prod.* 142:111877. doi: 10.1016/j.indcrop.2019.111877

Li, H., Zhang, L., Hu, J., Zhang, F., Chen, B., Xu, K., et al. (2017). Genome-wide association mapping reveals the genetic control underlying branch angle in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 8:1054. doi: 10.3389/fpls.2017.01054

Li, K., Wang, J., Kuang, L., Tian, Z., Wang, X., Dun, X., et al. (2021). Genome-wide association study and transcriptome analysis reveal key genes affecting root growth dynamics in rapeseed. *Biotechnol. Biofuels* 14:178. doi: 10.1186/s13068-021-02032-7

Li, N., Song, D., Peng, W., Zhan, J., Shi, J., Wang, X., et al. (2019). Maternal control of seed weight in rapeseed (*Brassica napus* L.): The causal link between the size of pod (mother, source) and seed (offspring, sink). *Plant Biotechnol. J.* 17, 736–749. doi: 10.1111/pbi.13011

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.* 15:01040. doi: 10.1104/pp.15.01040

Li, S., Zhu, Y., Varshney, R. K., Zhan, J., Zheng, X., Shi, J., et al. (2020). A systematic dissection of the mechanisms underlying the natural variation of silique number in rapeseed (*Brassica napus* L.) germplasm. *Plant Biotechnol. J.* 18, 568–580. doi: 10.1111/pbi.13224

Li, Y. P., Cheng, Y., Cai, G. Q., Fan, C. C., and Zhou, Y. M. (2014). Cytological basis and molecular mechanism of variation in number of seeds per pod in *Brassica napus*. *Sci. Sin. Vitae* 44, 822–831. doi: 10.1360/052014-87

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

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., et al. (2017). Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. *Front. Plant Sci.* 8:206. doi: 10.3389/fpls.2017.00206

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Nemhauser, J. L., Feldman, L. J., and Zambryski, P. C. (2000). Auxin and *ETTIN* in *Arabidopsis* gynoecium morphogenesis. *Development* 127:3877–3888. doi: 10.1242/dev.127.18.3877

Okada, K., Ueda, J., Komaki, M. K., Bell, C. J., and Shimura, Y. (1991). Requirement of the auxin polar transport system in early stages of *Arabidopsis* floral bud formation. *Plant Cell.* 7, 677–684. doi: 10.1105/tpc.3.7.677

Qadir, M., Wang, X., Shah, S. R. U., Zhou, X. R., Shi, J., and Wang, H. (2021). Molecular network for regulation of ovule number in plants. *Int. J. Mol. Sci.* 22:12965. doi: 10.3390/ijms222312965

Shi, J., Zhan, J., Yang, Y., Ye, J., Huang, S., Li, R., et al. (2015). Linkage and regional association analysis reveal two new tightly-linked major-QTLs for pod number and seed number per pod in rapeseed (*Brassica napus* L.). *Sci. Rep.* 5:14481. doi: 10.1038/srep14481

Strelin, M. M., and Aizen, M. A. (2018). The interplay between ovule number, pollination and resources as determinants of seed set in a modular plant. *PeerJ.* 6:e5384. doi: 10.7717/peerj.5384

Vera-Sirera, F., Gomez, M. D., and Perez-Amador, M. A. (2016). "DELLA proteins, a group of GRAS transcription regulators that mediate gibberellin signaling" in Plant transcription factors: Evolutionary, structural and functional aspects, Vol. 20, Ed. Sience Direct (Cambridge, MA: Academic Press), 313–328 [Sience Direct]. doi: 10.1016/B978-0-12-800854-6.00020-8

Wang, J.-G., Feng, C., Liu, H.-H., Ge, F.-R., Li, S., Li, H.-J., et al. (2016). *HAPLESS13*-mediated trafficking of *STRUBBELIG* is critical for ovule development in *Arabidopsis*. *PLoS Genet.* 12:e1006269. doi: 10.1371/journal.pgen.1006269

Wang, T., Wei, L., Wang, J., Xie, L., Li, Y. Y., Ran, S., et al. (2020). Integrating GWAS, linkage mapping and gene expression analyses reveals the genetic control of growth period traits in rapeseed (*Brassica napus* L.). *Biotechnol. Biofuels* 13, 134. doi: 10.1186/s13068-020-01774-0

Wetzstein, H. Y., Yi, W., Porter, J. A., and Ravid, N. (2013). Flower position and size impact ovule number per flower, fruitset, and fruit size in pomegranate. *J. Am. Soc. Hortic. Sci.* 138, 159–166. doi: 10.21273/jashs.138.3.159

Yang, X., and Tucker, M. R. (2021). Establishing a regulatory blueprint for ovule number and function during plant development. *Curr. Opin. Plant Biol.* 63:102095. doi: 10.1016/j.pbi.2021.102095

Yang, Y., Shen, Y., Li, S., Ge, X., and Li, Z. (2017a). High density linkage map construction and QTL detection for three silique-related traits in orychophragmus violaceus derived *Brassica napus* population. *Front. Plant Sci.* 8:1512. doi: 10.3389/fpls.2017.01512

Yang, Y., Wang, Y., Zhan, J., Shi, J., Wang, X., Liu, G., et al. (2017b). Genetic and cytological analyses of the natural variation of seed number per pod in rapeseed (*Brassica napus* l.). *Front. Plant Sci.* 8:1890. doi: 10.3389/fpls.2017.01890

Yu, S.-X., Zhou, L.-W., Hu, L.-Q., Jiang, Y.-T., Zhang, Y.-J., Feng, S.-L., et al. (2020). Asynchrony of ovule primordia initiation in *Arabidopsis*. *Development* 147:196618. doi: 10.1242/dev.196618

Yu, S.-X., Jiang, Y.-T., and Lin, W.-H. (2022). Ovule initiation: The essential step controlling offspring number in *Arabidopsis*. *J. Integr. Plant Biol.* 64, 1469–1486. doi: 10.1111/jipb.13314

Yuan, J., and Kessler, S. A. (2019). A genome-wide association study reveals a novel regulator of ovule number and fertility in *Arabidopsis thaliana*. *PLoS Genet.* 15:e1007934. doi: 10.1371/journal.pgen.1007934

Zhu, Y., Ye, J., Zhan, J., Zheng, X., Zhang, J., Shi, J., et al. (2020). Validation and characterization of a seed number per silique quantitative trait locus *qSN.A7* in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 11:68. doi: 10.3389/fpls.2020.00068

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fpls.2021.780250

Exploiting High-Throughput Indoor Phenotyping to Characterize the Founders of a Structured *B. napus* Breeding Population

Jana Ebersbach[1†], Nazifa Azam Khan[2†], Ian McQuillan[2], Erin E. Higgins[1], Kyla Horner[1], Venkat Bandi[2], Carl Gutwin[2], Sally Lynne Vail[1], Steve J. Robinson[1] and Isobel A. P. Parkin[1]*

[1] Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, [2] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:
Lewis Lukens,
University of Guelph, Canada

Reviewed by:
Ravi Valluru,
University of Lincoln, United Kingdom
John Sulik,
University of Guelph, Canada

***Correspondence:**
Isobel A. P. Parkin
isobel.parkin@agr.gc.ca

[†] These authors have contributed equally to this work

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

Received: 20 September 2021
Accepted: 10 December 2021
Published: 05 January 2022

Citation:
Ebersbach J, Khan NA, McQuillan I, Higgins EE, Horner K, Bandi V, Gutwin C, Vail SL, Robinson SJ and Parkin IAP (2022) Exploiting High-Throughput Indoor Phenotyping to Characterize the Founders of a Structured B. napus Breeding Population. Front. Plant Sci. 12:780250. doi: 10.3389/fpls.2021.780250

Phenotyping is considered a significant bottleneck impeding fast and efficient crop improvement. Similar to many crops, *Brassica napus*, an internationally important oilseed crop, suffers from low genetic diversity, and will require exploitation of diverse genetic resources to develop locally adapted, high yielding and stress resistant cultivars. A pilot study was completed to assess the feasibility of using indoor high-throughput phenotyping (HTP), semi-automated image processing, and machine learning to capture the phenotypic diversity of agronomically important traits in a diverse *B. napus* breeding population, SKBnNAM, introduced here for the first time. The experiment comprised 50 spring-type *B. napus* lines, grown and phenotyped in six replicates under two treatment conditions (control and drought) over 38 days in a LemnaTec Scanalyzer 3D facility. Growth traits including plant height, width, projected leaf area, and estimated biovolume were extracted and derived through processing of RGB and NIR images. Anthesis was automatically and accurately scored (97% accuracy) and the number of flowers per plant and day was approximated alongside relevant canopy traits (width, angle). Further, supervised machine learning was used to predict the total number of raceme branches from flower attributes with 91% accuracy (linear regression and Huber regression algorithms) and to identify mild drought stress, a complex trait which typically has to be empirically scored (0.85 area under the receiver operating characteristic curve, random forest classifier algorithm). The study demonstrates the potential of HTP, image processing and computer vision for effective characterization of agronomic trait diversity in *B. napus,* although limitations of the platform did create significant variation that limited the utility of the data. However, the results underscore the value of machine learning for phenotyping studies, particularly for complex traits such as drought stress resistance.

Keywords: *B. napus*, spring-type, NAM, semi-automated image analysis, machine learning, drought resistance

INTRODUCTION

Ensuring and increasing food security for a growing global population faced with uncertain environmental changes is one of the main challenges of agricultural research in the 21st century (Tilman et al., 2011; Hickey et al., 2019; Lenaerts et al., 2019). In addition to increasing the productivity of current arable land, it will be crucial to increase crop yield to meet rising global demand; however, yield improvement is currently progressing at an insufficient pace (Tilman et al., 2011; Ray et al., 2013). Even though technological advances, such as genomic selection and doubled haploid technology, have resulted in substantial acceleration of the breeding process, development of new high yield, pest and disease resistant, and climate-smart crop varieties is still hampered by several factors. These include long generation times and time-consuming steps such as phenotypic evaluation of large populations (Hickey et al., 2019). Accurate phenotyping, in particular, is considered one of the major bottlenecks of modern crop breeding, which has led to a strong emphasis on the development of automated, scalable, non-destructive, and high-throughput imaging approaches (High-Throughput Phenotyping, HTP). HTP could augment traditional methods of phenotypic trait quantification that are time and labor intensive, subjective, and often destructive. Over recent years, HTP applications have seen dramatic advances afforded by the continual improvement to sensor and automation technologies (Yang et al., 2020). In addition to corresponding advances in data acquisition, one of the biggest innovations in the field of HTP is the use of machine learning for automated analysis of the vast amounts of data generated by HTP platforms (Zhao et al., 2019). However, application of machine learning to the problem of phenotyping is still in its infancy and widespread deployment of these tools will require further refinements to algorithms and analysis pipelines.

Many pressing challenges with regards to crop phenotyping remain, including the measurement of multigenic or multidimensional traits and the dissection of complex phenotypes that are hard to reliably reproduce in field settings, such as abiotic stress responses, which are often the culmination of multiple environmental factors acting simultaneously (Yang et al., 2020). Some of these challenges can be overcome using indoor phenotyping systems, where growing conditions and imaging parameters can be controlled more precisely allowing individual plant architectural and physiological traits to be measured with greater accuracy. This renders controlled environment phenotyping particularly suitable for forward and reverse genetics, as well as quantitative genetics and genetic mapping (Mir et al., 2019; Yang et al., 2020). However, due to the limited space available in the small number of existing large-scale indoor phenotyping facilities as well as the high costs associated with running these experiments, only a handful of annotated benchmark image datasets are currently publicly available, mostly in Arabidopsis and grain crops (e.g., Fahlgren et al., 2015; Choudhury et al., 2016; Cruz et al., 2016; Minervini et al., 2016; Veley et al., 2017[1,2]).

Brassica napus is a multipurpose crop of major economic importance, especially the oilseed morphotype, which is a source of vegetable oil for human consumption, industrial feedstock, and protein rich meal used in animal feed (Friedt et al., 2018). Due to its relatively recent allotetraploid origin, *B. napus* has a narrow genetic base which was further eroded during initial domestication and extensive breeding activities throughout the last century (Diers and Osborn, 1994; Becker et al., 1995; Rahman, 2013; Gazave et al., 2016). Similar to other crops, this erosion needs to be addressed and further crop improvement will require the introduction of new genetic variation into current elite cultivars (Rahman, 2013; Friedt et al., 2018; Rebetzke et al., 2019). Among other approaches, this will necessitate the systematic screening of diverse germplasm collections for desirable phenotypic traits as well as extensive pre-breeding activities. In addition, dissection of the genetic basis for key traits targeted in rapeseed breeding is considered crucial for accelerated crop development (Knoch et al., 2020). Development of automated HTP protocols optimized for *B. napus*, paired with effective genomic trait dissection strategies, holds great promise to accelerate the achievement of breeding targets, such as short crop cycles, high yield, and resistance to heat and drought conditions (Delourme et al., 2018). However, comprehensive, publicly available indoor HTP datasets of *B. napus* are currently limited to 2D root phenotyping (Thomas et al., 2016a,b; Zhang et al., 2016) and early stage phenotyping (Kjaer and Ottosen, 2015; Pommerrenig et al., 2018; Knoch et al., 2020), thus the full potential of HTP technologies for above ground trait quantification has yet to be explored.

Here, we used a diverse panel of 51 spring-type *B. napus* lines, selected as founders for the development of a large, spring-type *B. napus* Nested Association Mapping (NAM) population called "SKBnNAM," to test the feasibility of using current indoor HTP technology for rapid, semi-automated phenotyping of several key traits, including flowering and canopy architecture traits (timing of anthesis, number of flowers, and number of raceme branches) and resistance to drought. NAM is a powerful trait dissection strategy that combines association and linkage genetic mapping with genome sequencing, and it has been used to elucidate the underlying genetic architecture of several important agronomic traits in other crops, such as maize (Yu et al., 2008), barley (Maurer et al., 2015), and soybean (Song et al., 2017). The work provides a comprehensive reference image dataset for *B. napus*, as well as computational methods that can be used to extract maximal information from such datasets, including a novel method for prediction of drought stress. The work also exposes some limitations of current HTP platforms, which should be considered when considering such an approach.

MATERIALS AND METHODS

NAM Founder Selection and Genotyping

A total of 297 *B. napus* lines from global breeding collections were genotyped using the *Brassica* 60K Illumina Infinium array as described in Clarke et al. (2016) (**Supplementary Table 3.1**). The genotype data was visualized using the GenomeStudio software suite (Illumina, Inc.) and a custom cluster file developed for

[1]https://plantvision.unl.edu/dataset

[2]https://www.quantitative-plant.org/dataset

B. napus was applied to screen out poorly performing and multi-locus SNPs as described in Clarke et al. (2016). The data was filtered for a minor allele frequency of 0.05 and a minimum separation of the homozygous A and B clusters of 0.8 to eliminate monomorphic and multi-locus SNPs resulting in a total of 30,933 SNPs for 297 accessions. PCA was carried out using the R package SNPRelate (Zheng et al., 2012). Vcftools v.0.1.16 (Danecek et al., 2011) was used for additional data filtering (maximum missing proportion of 0.8).

A total of 51 spring-type (or annual) lines were selected from the larger dataset to establish a *Brassica napus* NAM population (SKBnNAM) (**Supplementary Table 3.2**). Founder line selection was based on an assessment of levels of heterozygosity, relatedness, available phenotype data, as well as geographic origin. The chosen founder lines encompassed material from countries where spring rapeseed production is an economic priority and these were augmented with material from a diverse geography as well as the inclusion of synthetic rapeseed. The genotype data allowed the selection of the most diverse array of inbred material. The line N99-508 (NAM 0) was selected as the common parent for the NAM population. N99-508 is adapted to the Canadian environment, which potentiates the evaluation of the NAM population under Canadian field conditions.

LemnaTec Dataset

Plants were grown in the LemnaTec Scanalyzer 3D facility at University of Nebraska, Lincoln campus. In total, 50 out of 51 NAM founder lines were grown in six replicates, as NAM 10 failed to germinate. All plants were sown, and grown at 18–22°C for 20 days (for detailed growing conditions, see **Supplementary Table 3.3**) before being loaded onto the phenotyping platform. From 21 days after seeding (DAS) to 34 DAS, plants were weighed before and after watering, and imaged once a day. During the treatment phase (35–55 DAS), three plants each were subjected to one of two different watering regimes (A: control, 100% field capacity and B: drought treatment at 40% field capacity) and imaged and weighed every day for the first 3 days, then every other day (**Figure 1**). During the post-treatment phase (57–67 DAS), all plants were watered equally and imaged and weighed every other day. At each phenotyping time point, images were captured for the following camera types and angles: Visible light (RGB) from 10 angles, Infrared from 6 angles, Near Infrared (NIR) from top view, fluorescent light (FLUOR) from 6 angles and hyperspectral wavelengths (HYP) from one angle. The full image dataset is openly available at https://p2irc-data-dev.usask.ca/dataset/10.1109.SciDataManager.2020.7284788 (Dataset name: P2IRC Flagship 1 Data). Camera specifications and more detailed descriptions of the phenotyping facility are available in Choudhury et al. (2016, 2018).

Dataset annotations (ground truths) were generated via manual evaluation and scoring of selected traits in both treatment groups. First, all images were assessed for presence of open flowers (0 = not flowering, 1 = flowering), with the earliest date

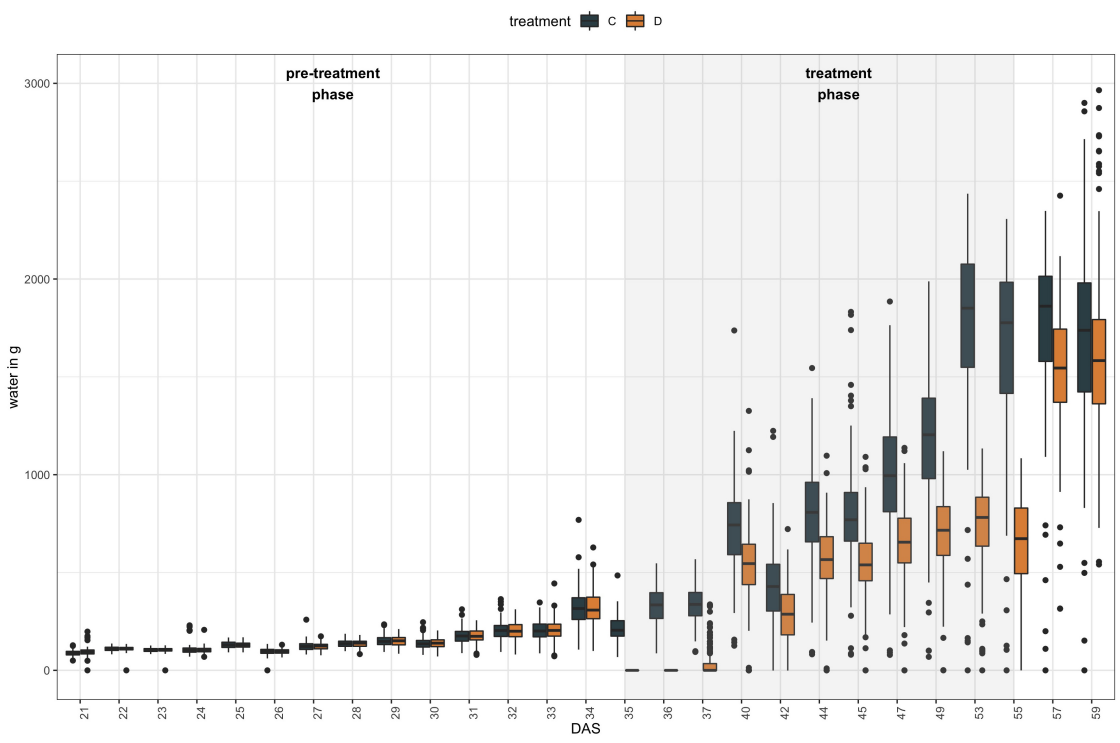


FIGURE 1 | Experimental set-up of high-throughput indoor phenotyping of 50 *B. napus* founder lines at LemnaTec Scanalyzer 3D facility at University of Nebraska. Six replicates were grown of each genotype and three plants were subjected to one of two watering regimes between 35 DAS and 55 DAS: control plants (C) were kept at 100% field capacity (blue bars), while drought treated (D) plants were maintained at 40% field capacity during that period (orange bars).

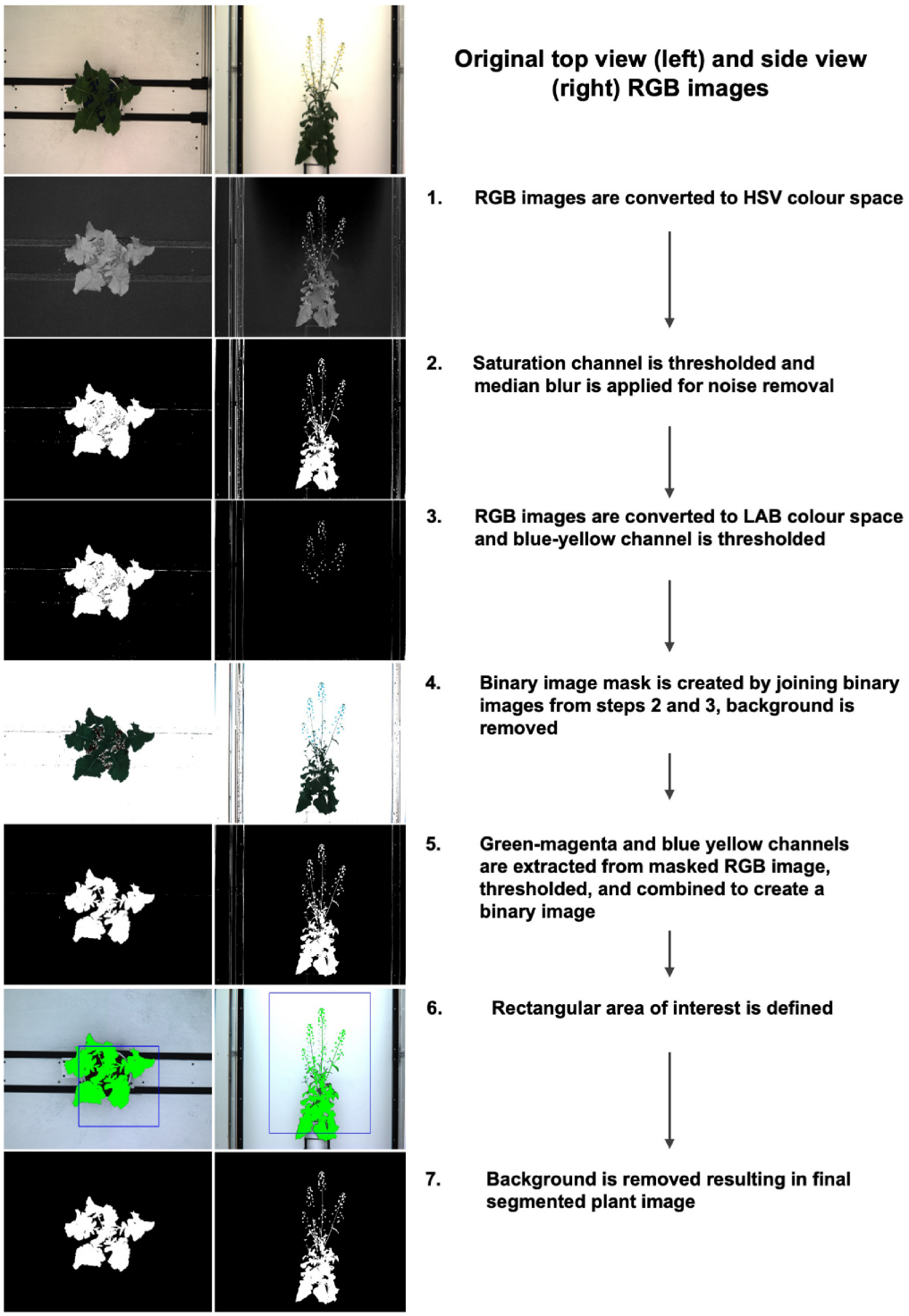


FIGURE 2 | Image segmentation workflow and corresponding examples of top view (left) and side view (right) images.

of open flowers being counted as the date of anthesis. Then, the number of raceme branches was counted for each plant and phenotyping time point. Finally, plants were assessed for symptoms of drought stress using both top view and side view RGB images, using a binary scoring scheme in which healthy looking plants were scored as unstressed (0) and any plant exhibiting mild to severe drought stress (e.g., drooping leaves) was scored as stressed (1). Due to considerable time investment

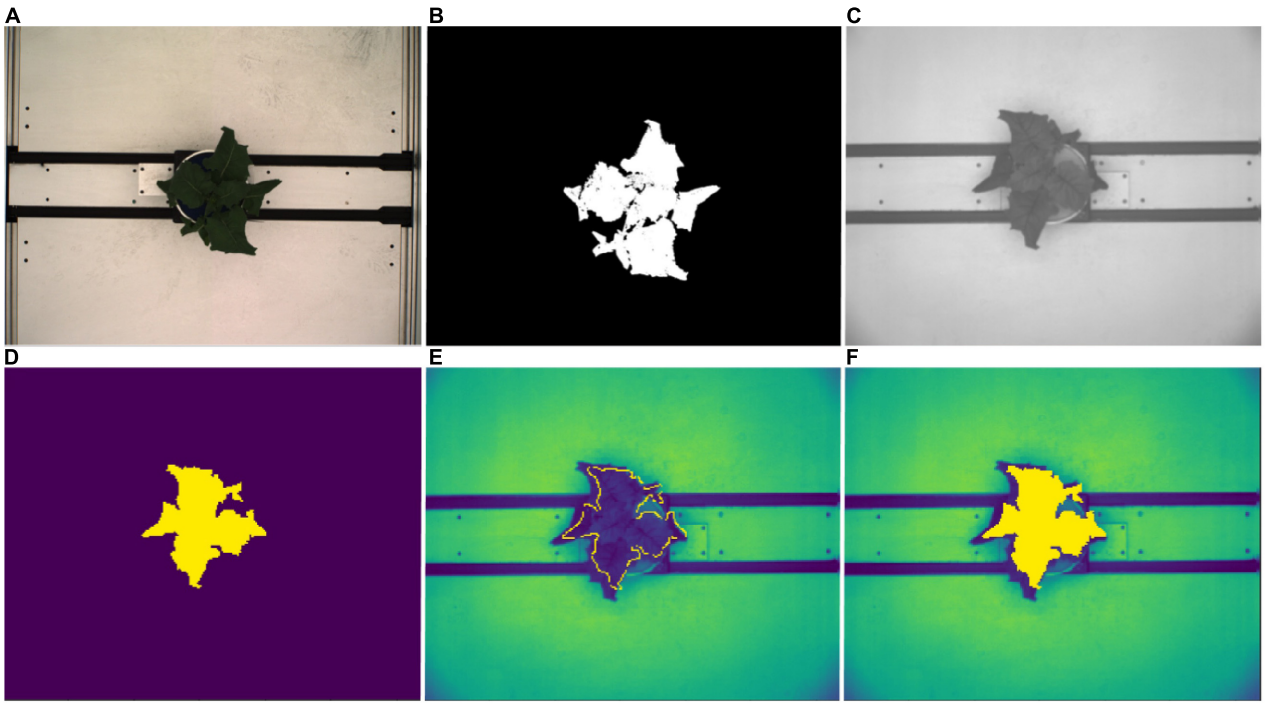


FIGURE 3 | NIR image processing workflow. **(A)** Top view RGB image. **(B)** Corresponding image segmentation resulting from workflow described in **Figure 2**. **(C)** Original top view NIR grayscale image. **(D)** Resized and rescaled RGB image segmentation used as an image mask for NIR image. **(E)** Segmented plant outline overlaying NIR image (yellow boundary). **(F)** Segmented plant area overlaying NIR image (yellow region).

necessary to score this trait as well as the exploratory nature of this study, this ground truth was restricted to 49 DAS, which was toward the end of the experimental treatment phase but before the majority of plants had started to flower.

Image Processing

For the purpose of this study, we concentrated on processing the RGB images using top view (90°) and side view (0°), and the NIR images (top view, 90°) between 21 DAS and 57 DAS. In order to remove as much background as possible, RGB images were segmented in Python 3 using PlantCV v.2 (Gehan et al., 2017) according to the steps outlined in the program documentation (**Figure 2**, see **Supplementary Appendix 1** for extended methods). In short, this procedure involves converting the RGB image to the HSV (Hue, Saturation, and Value) and the CIELAB color space, followed by several thresholding steps, and the creation of a binary mask which is then applied to the original image to retain only plant-related information. Although manual identification of a range of thresholds made the segmentation process semi-automated, the lack of available ground truth data was the primary reason for not using automated segmentation methods with machine learning approaches. There were 300 plants and ca. 28 imaging days including both side and top view images for a total of 15988 images. Annotating these images would have been significantly more time consuming than finding the appropriate set of thresholds. Note that the segmentations were evaluated visually, by making graphs of extracted convex-hull of the plant areas. The graphs and images of 300 plants were visually checked to iteratively improve segmentation and finalize threshold values. Following image segmentation, several basic, holistic phenotypic values were extracted from each image which allowed for a numeric characterization of each plant at each phenotyping time point: the convex hull area of each plant from both top view and side view, the number of plant pixels (projected leaf area) from both top view and side view, and the plant height and width. In addition, in order to track minor leaf color changes, which could be indicative of drought stress, the Excess Green Index (ExG = 2G – R + B, Woebbecke et al., 1995) was extracted from the segmented RGB images.

Leaf spectral reflectance is known to be substantially driven by leaf water content and NIR wavelengths have been shown to be useful in tracking plant water status (Berger et al., 2010; Briglia et al., 2019). Thus image masks from the RGB top view segmentations were fitted to the NIR images in order to extract the plant pixels of the grayscale NIR images (**Figure 3**). Following Vello et al. (2015) and Janni et al. (2019), NIR pixel intensities were summarized using mean and 75th percentile NIR values for each plant and day of imaging. However, due to divergent fields-of-view of the RGB and NIR cameras, this was only possible between 37 DAS and 59 DAS.

The plant boundary curvatures were extracted from the segmented top view images using the histogram of curvature over scale method (HOCS; Kumar et al., 2012). This robust, and rotation-invariant shape descriptor computes the curvature at each point on the region boundary (plant boundary) at a range of different scales, and then quantizes curvature via histograms

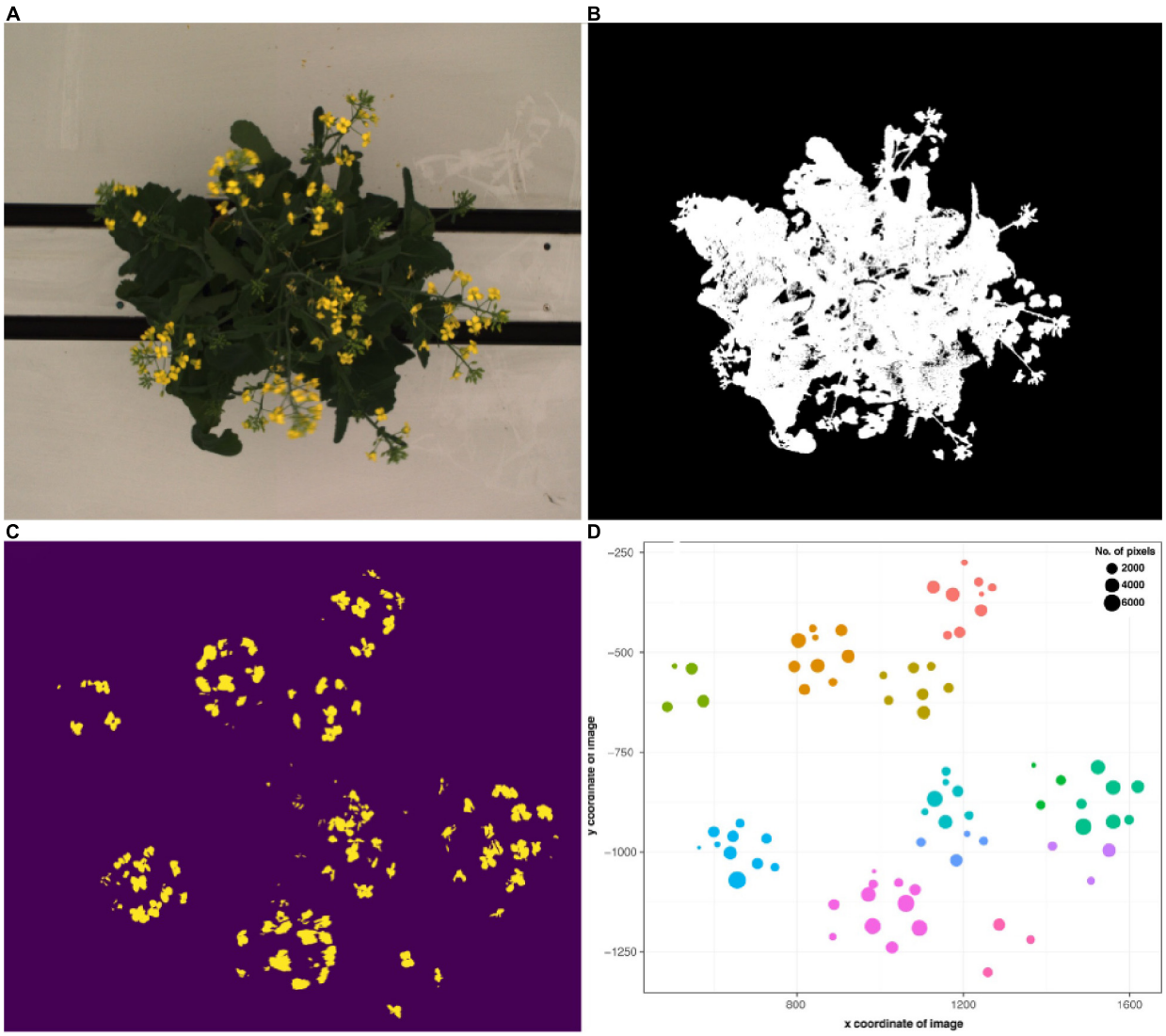


FIGURE 4 | Flower and raceme branch detection workflow. **(A)** Exemplary RGB top view image of a flowering *B. napus* individual. **(B)** Segmented image. **(C)** Image components in the blue yellow channel after pixel-based thresholding. **(D)** Binary image components mapped according to their x and y coordinates and number of pixels. Colors correspond to clusters assigned via hierarchical clustering and roughly correspond to raceme branches.

(Van Vliet and Verbeek, 1993). Boundary curvatures on 49 DAS were computed over the normalized area integral invariant (Manay et al., 2006) at 25 different circle radii (scales) ranging from 5 to 125 (increment of 5) and summarized as histograms of curvature over scale using 5 bins per histogram at each scale.

Semi-automated flower detection and quantification was performed by applying the segmented RGB top view image as a mask to the blue yellow channel of the converted LAB color space top view image (**Figure 4**). Then, pixel-based thresholding was performed according to plant age and treatment group, and the number of pixels ("flower pixels") in the thresholded blue yellow channel image was extracted (see **Supplementary Appendix 1** for full details). The components (i.e., pixel clusters) of the resulting binary image were considered to be either single, or several overlapping flowers (**Figure 4C**) and their descriptive properties (number of component pixels, component centroid coordinates, minimum and maximum x and y coordinates of bounding box around each component, area of each bounding box, convex hull area) were used to plot the flowers (**Figure 4D**) and derive several additional properties (e.g., maximum canopy width and canopy angle). The predicted flower annotations were then compared to manual flower scores as described above.

Data Analysis

All image processing data was analyzed under the tidyverse framework in R v. 4.0.2 (R Core Team, 2020) and RStudio v. 1.2.1335. Besides manually removing all data for plants which did not germinate, a threshold-based outlier removal procedure (median ± 3 standard deviations) was applied for each trait and

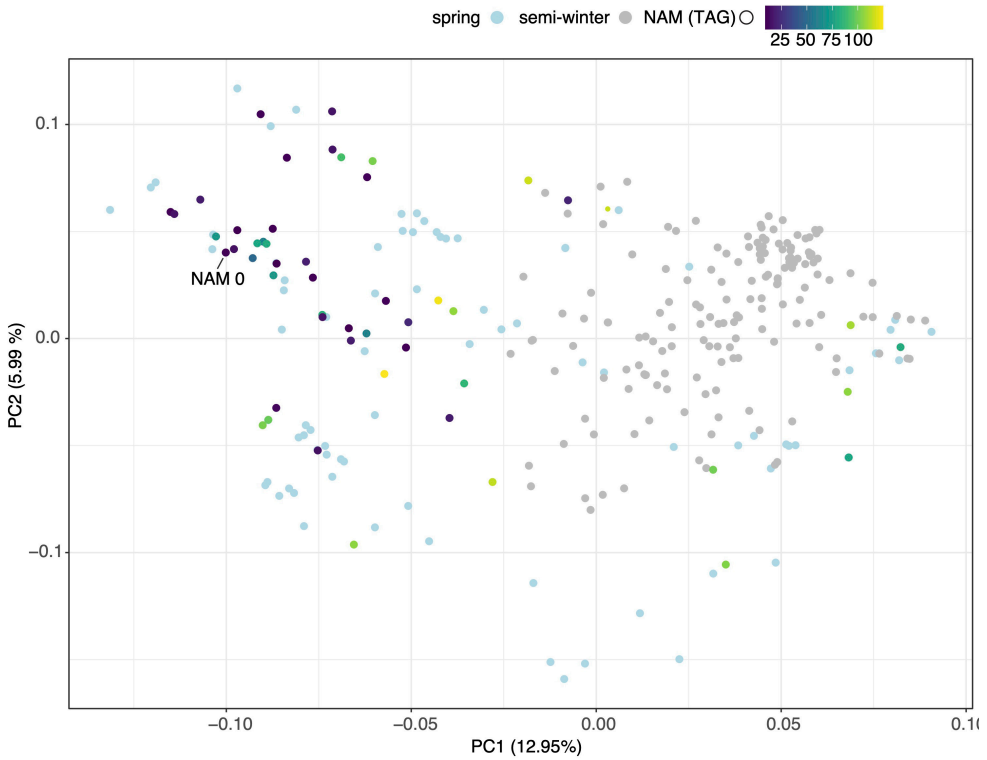


FIGURE 5 | Principal Component Analysis of genotypic diversity of 297 spring (light blue dots) and semi-winter (gray dots) *B. napus* varieties. Founder lines of SKBnNAM introduced in this study are colored according to their mean total aliphatic glucosinolate (TAG) levels.

day, separately (Knoch et al., 2020). In order to track day-to-day changes in plant development, day-to-day differences were calculated for most extracted traits of interest (denoted by Δ). Further, image-based phenotypes were used to derive estimated biovolume [$\sqrt{}$(side projected leaf area^2 * top projected leaf area); Junker et al., 2015], which has previously been shown to be a suitable proxy for plant biomass in canola (Knoch et al., 2020). In addition, the results from the flower detection protocol (see above) were used to extract the maximum canopy width and the approximate canopy angle in order to identify genotypes with particularly loose or particularly compact inflorescences, a common consideration for breeders.

The number of components in the blue-yellow channel, excluding those of less than 100 pixels, was used to estimate the number of flowers per plant and imaging day. Large components, that likely included several overlapping flowers, were split by dividing their total number of pixels by the approximate number of pixels of a fully opened flower viewed from the top. Hierarchical clustering was applied using the hclust function in R using the extracted coordinates of the filtered flower components (min x, max x, min y, max y, centroid x, and centroid y) followed by pruning the resulting hierarchical tree at the total number of manually assessed raceme branches (i.e., groups, k). Spearman correlation coefficients and significance levels for selected traits were calculated in R using cor.test. Processed trait data was visualized using the R package ggplot2 (Wickham, 2016).

Machine Learning

Several supervised machine learning models were tested for prediction and identification of two key agronomical traits: the number of raceme branches and drought stress. The manually assessed ground truth of the number of raceme branches consisted of 789 observations from 39 genotypes that started flowering within the experimental period with 34 different class labels (number of raceme branches). Since raceme branch counts were unevenly distributed [e.g., 198 of 789 images were of flowering individuals with one inflorescence branch (main stem) while only 21 images featured plants with 10 raceme branches], this was considered a regression rather than a classification problem. The data set was randomly divided into 5 sets for a 5-fold cross validation, with 4 sets serving as training data and one set serving as testing data in 5 replicate experiments. For each set, plant age in DAS, the filtered number of pixels in the blue-yellow channel (i.e., "flower pixels") and the estimated number of flowers were used as input features for raceme branch number prediction. The widely used linear regression and Huber regression algorithms were applied to predict the inflorescence branch numbers (Huber, 1992; Freedman, 2009). Prediction accuracy of different machine learning algorithms was then gauged by comparing to the manually established ground truth of raceme branch numbers.

With regard to drought stress, the number of stressed plants (n = 166) was randomly down-sampled to match the number

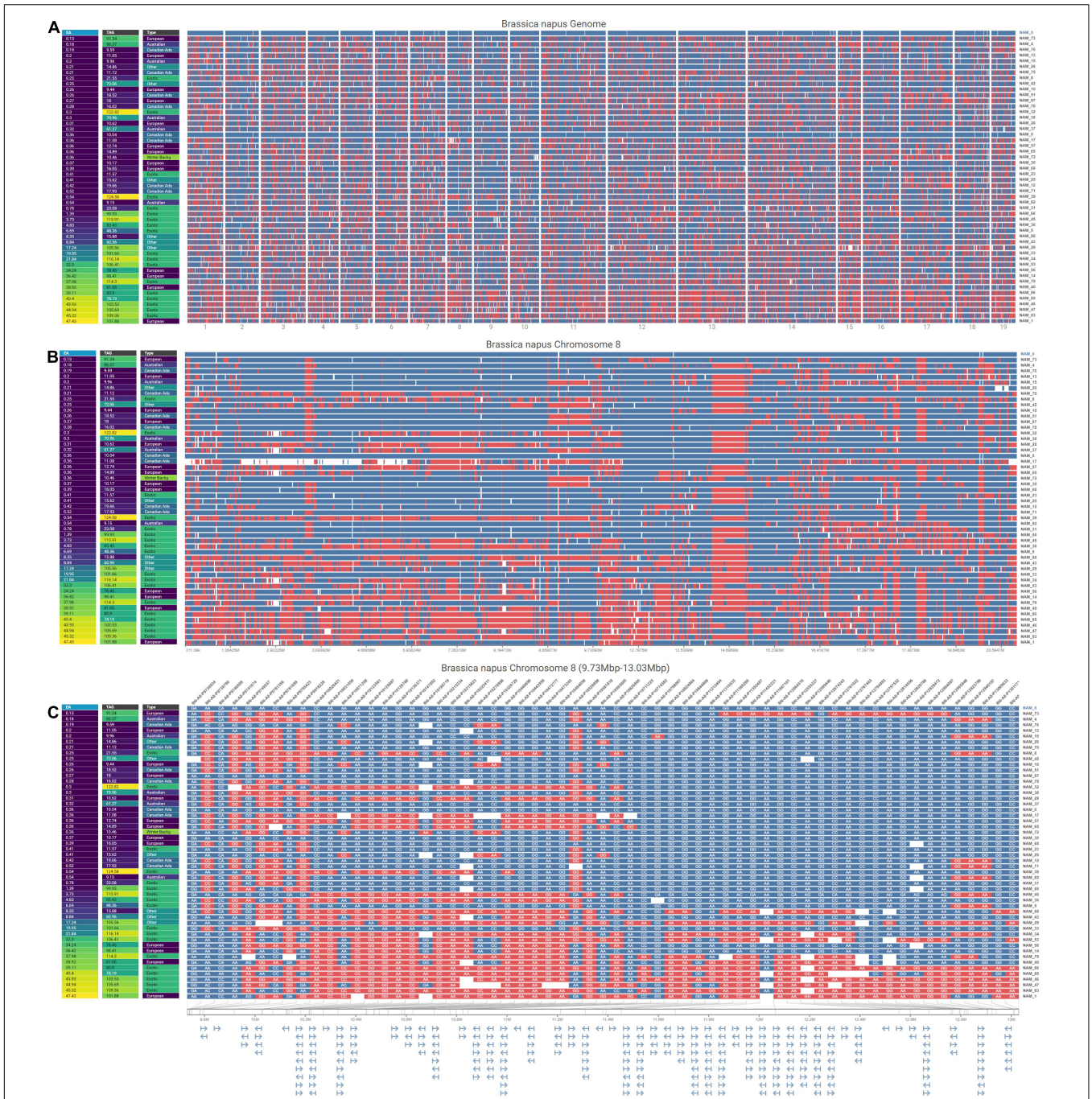


FIGURE 6 | Visualizations at three resolutions comparing SNPs in 50 *B. napus* genotypes against a reference genotype (NAM 0) with the genotypes ordered by their increasing erucic acid content. The visualizations are all organized with rows representing different lines, and columns of pixels representing each SNP. The color at each SNP indicates whether that SNP is the same (blue) as or different (red) from the allele in the reference genotype shown at the top of the view and missing data is encoded in white. Additionally, at each resolution, a map of phenotypic trait values (erucic acid content, total aliphatic glucosinolate levels, and genotype origin) are shown on the left. **(A)** Genome-level overview of the entire *B. napus* genome horizontally separated by chromosome. **(B)** Chromosome-level view of chromosome 8 in *B. napus*. **(C)** SNP-level view of chromosome 8 (9.73 mb to 13.03 mb) that displays SNP names at the top of the column, nucleotides for each SNP, and a gene map below the main view.

of unstressed plants (n = 133), resulting in a total dataset of 266 observations. Combinations of several input features were used to test machine learning for stress detection, including mean and 75th percentile NIR, difference in plant pixels between 47 DAS and 49 DAS (Δ pixelsTV, assuming that drooping leaves might result in a reduction in plant size as seen from above),

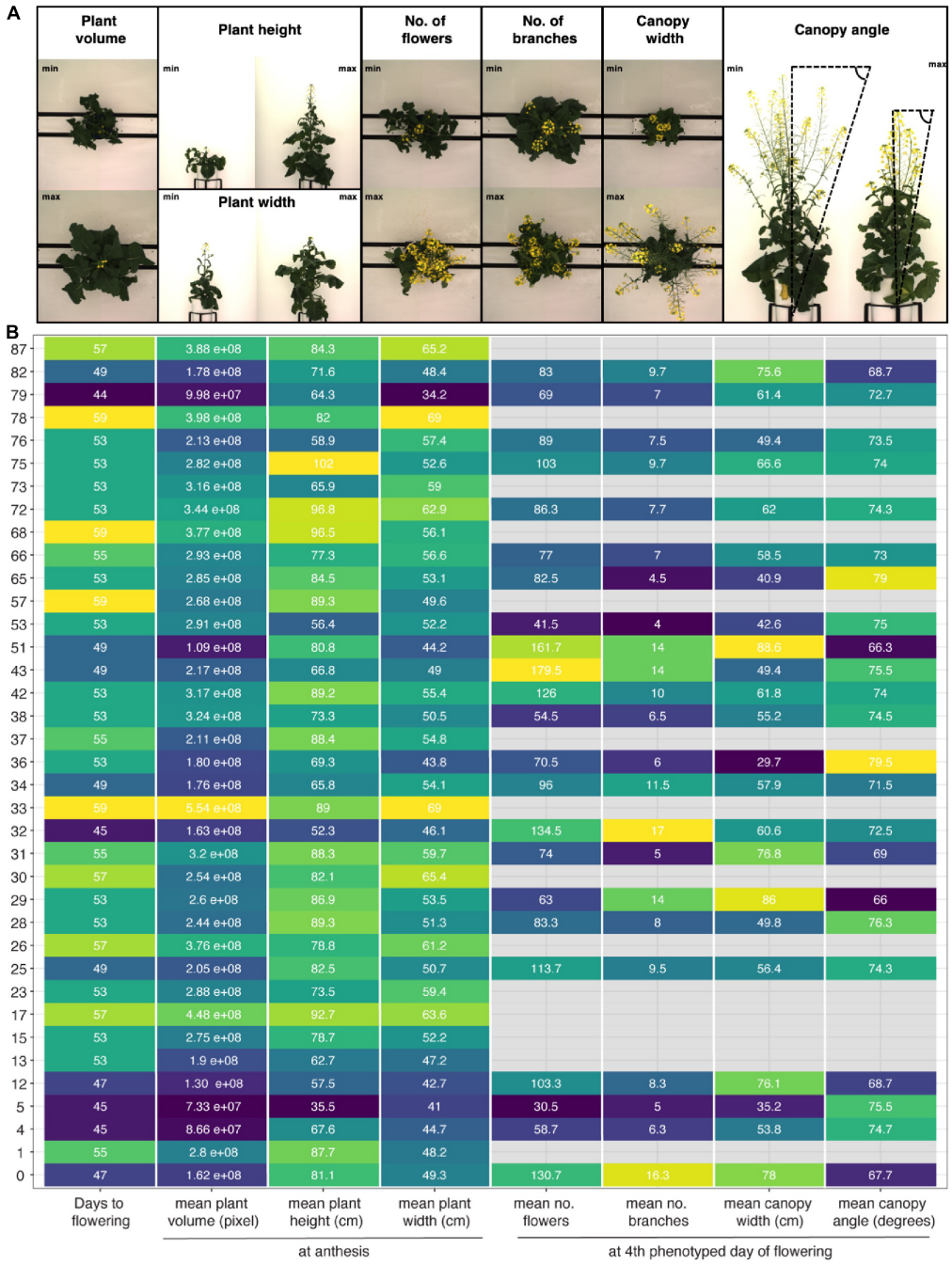


FIGURE 7 | Morphological diversity among founder lines of SKBnNAM based on eight selected growth traits extracted from the dataset. **(A)** Images show individuals representing minimum and maximum values for each growth trait (except for days to flowering). **(B)** Mean values for each trait and genotype (control group only). Tiles are colored based on normalized trait means (darker colors indicating lower values, lighter colors indicating higher values). Absolute mean values are given in each tile. Only genotypes that started to flower before DAS 59 are included in this plot (n = 37).

the convex hull area (top view), total number of plant pixels (top view), Δ convex hull (top view), plant height and width, Δ plant height and Δ plant width, ExG, and number of raceme branches. Again, the data set was randomly divided into 5 subsets for a 5-fold cross validation, with 4 sets serving as training data and one set serving as testing data in 5 replicate experiments. Stress identification was tested with the commonly used supervised machine learning algorithms Random Forest Classifier, Linear Discriminant Analysis, Logistic Regression model, K-nearest Neighbor, Decision Tree, and Support Vector Machine, most of which had been previously employed for stress detection in plants (Singh et al., 2016).

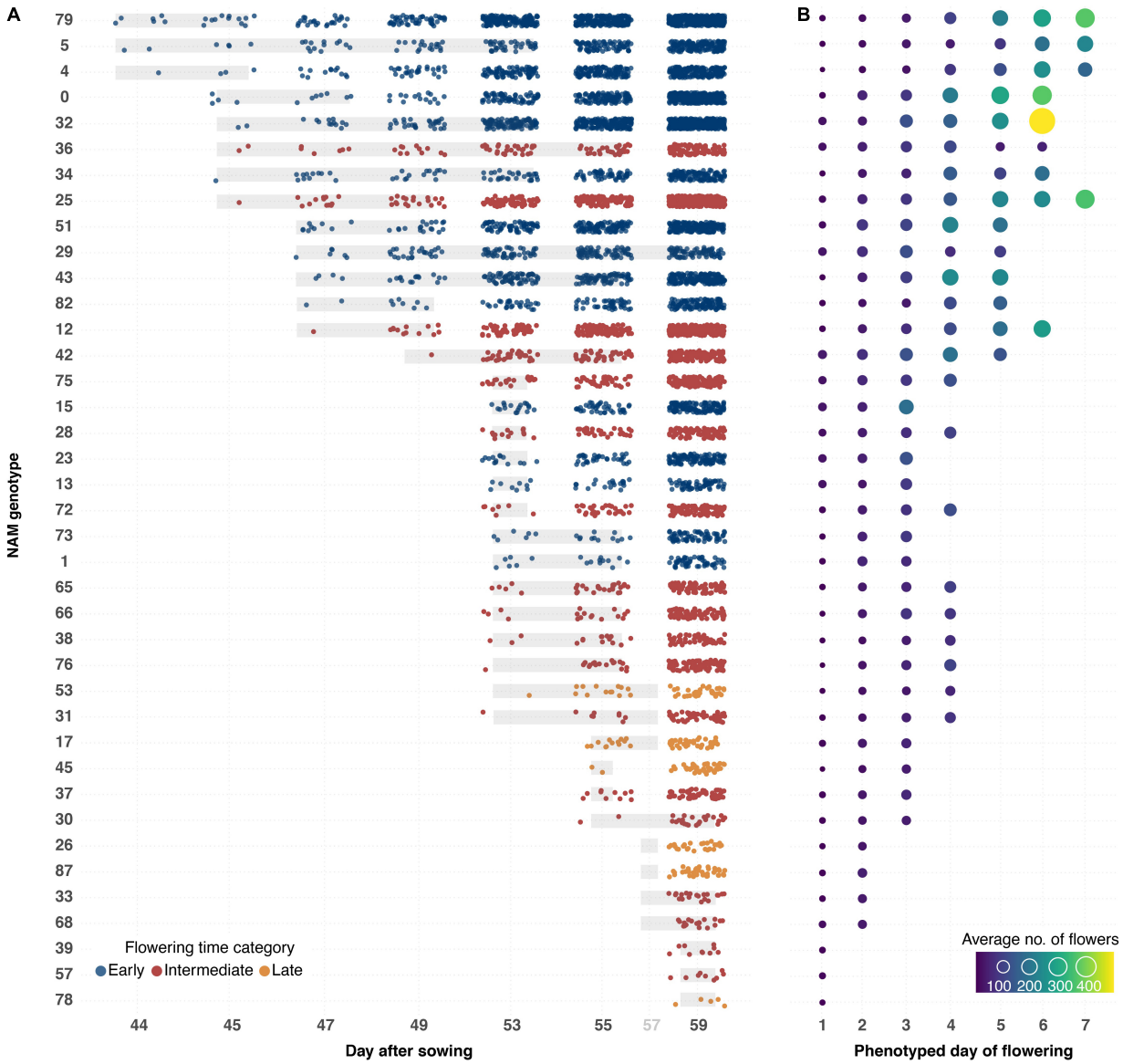


FIGURE 8 | Flowering phenology and intensity for 50 *B. napus* genotypes ranked by date of anthesis and flower output. **(A)** Flowering onset and intensity of *B. napus* genotypes that flowered during the LemnaTec experiment. Dots represent average estimated number of flowers recorded from three individuals (control group) of each genotype. The dots are colored according to previous greenhouse observations of flowering time. Underlying gray bars represent variation in flowering onset. **(B)** Flowering intensity per phenotyped day of flowering. Sizes and colors of circles represent average number of flowers recorded from three individuals (control group) of each genotype. Order of genotypes is shown in left-hand column in **(A)**.

RESULTS

The *Brassica napus* Germplasm

A total of 30,933 single nucleotide polymorphism (SNP) markers assayed from the Illumina *Brassica* Infinium SNP array were used to evaluate the genetic diversity among a wide collection of *Brassica napus* lines (**Supplementary Table 3.1**). The first two principal components of a principal component analysis (PCA) of 297 spring and semi-winter lines jointly explained 18.94% of the genetic variance and, similar to other studies, showed that these are positioned along a genetic gradient rather than constituting discriminate clusters (Lu et al., 2019). The founder lines of the SKBnNAM were selected to be well distributed among the genotyped spring-type accessions to capture most of the available variation (**Figure 5**). The available pedigrees and a summary of key seed quality traits (aliphatic glucosinolate and erucic acid content) segregating among the founders are provided in **Supplementary Table 3.2**. The SNP data for the founder lines has been represented in a novel graphical visualization tool[3] that allows relationships between the lines to be interrogated at the

[3]https://genomevis.usask.ca/haplotype-map-tree

genome, chromosome and SNP level (**Figure 6**). The lines can be arranged according to user-specified trait data in order to facilitate exploration of underlying genetic variation.

Plant Image Capture, Segmentation, and Trait Extraction

The initial HTP experiment intended to capture variation of the founder lines of the SKBnNAM in response to low water availability. Six replicates of each line were grown and imaged in a LemnaTec Scanalyzer 3D plant phenotyping facility (for an overview of growing conditions, see **Supplementary Table 3.3**). Thirty-four days after seeding (DAS) three of the replicates were maintained at field capacity (control group), while three were subjected to reduced watering to approximate 40% field capacity (treatment group). Imaging was carried out every day initially and every second day after DAS 37, and RGB images (top view, 90° and side view, 0°) as well as near-infrared (NIR) images were processed for specific phases of the experiment (see Materials and Methods). There was high variation for all extracted plant traits within genotypes and treatment groups (**Supplementary Figure 3.1**) and plants of both treatment groups displayed symptoms of drought stress (**Supplementary Figures 3.2,3.3**). This was likely caused by uneven watering administered during plant growth, mainly resulting from the large basal leaves blocking the automated watering system. Due to this variation, data analyses mainly focused on assessing and improving the efficacy of semi-automated image analysis protocols. Whenever genotype-specific traits were extracted (e.g., flowering time and canopy traits), analysis was limited to the control group.

Despite these challenges, the described image analysis pipeline allowed for extraction of holistic and derived phenotypic traits, including plant height, plant width and plant volume, a proxy for biomass, as well as plant phenology and canopy architecture traits (**Figure 7** and **Supplementary Figure 3.4A**, **Supplementary Appendix 2**). The method efficiently captured the wide-ranging diversity of growth and flowering traits among 50 out of 51 SKBnNAM founder lines (**Figure 7**), thereby demonstrating the potential of the population. While the exact thresholds used in the image segmentation and trait extraction process (**Supplementary Appendix 1**) are specific to this particular experiment, the outlined process used to derive these thresholds can be applied for other crops and experimental setups.

Flower Detection and Quantification

A semi-automated pipeline for flower detection was designed that allowed flowering to be accurately and precisely tracked throughout the experiment (**Figures 4**, **8**). This pipeline involved using the segmented top view RGB image as an image mask on the blue-yellow channel image of the same view, obtained by converting the RGB image to the LAB color space. Results of automatic flower detection were compared to manually determined dates of anthesis for each plant. After identification of the appropriate thresholds (see **Supplementary Appendix 1**), anthesis was correctly detected in 170 out of 175 flowering plants (97.14% accuracy). In four of the five cases where detection was not accurate, flower buds were observed in which the yellow petals were protruding from the sepals (**Supplementary Figure 3.5A**). The last case was caused by a single flower petal that had fallen from another plant (**Supplementary Figure 3.5C**). Following the successful detection of anthesis, the number of pixels as well as the number of components in the segmented image of the blue yellow channel (**Figure 4C**) were used to estimate the number of flowers for each plant and phenotyping time point. This approach allowed the flowering period of each plant to be measured through time, and for the different genotypes to be compared in terms of their phenology, flowering intensity, and variability among these characteristics (**Figure 8**). Early flowering varieties were clearly identified and the flowering behavior of these genotypes was consistent with those measured from previous field characterization (Spearman's rho = 0.76, **Supplementary Table 3.2**).

In order to facilitate the description of canopy architecture, quantification of raceme branches was explored by using the output from the flower detection pipeline as input to supervised machine learning algorithms; the results were then compared against manually counted raceme branch numbers for every plant at each phenotyped flowering time point. It was found that raceme branch numbers could be predicted from the estimated number of flowers and the total number of flower pixels with an accuracy of 91% ± 0.02 with both tested machine learning algorithms (Linear regression and Huber regression, both with 5-fold cross validation, **Table 1**).

TABLE 1 | Mean accuracies for inflorescence branch number prediction from image-derived traits for 50 *B. napus* genotypes using different machine learning algorithms.

	Raceme branch number prediction		
Number of observations (total/training data/testing data)	789/631/158		
	Mean accuracy with 5-fold cross validation (SD)		
	All features	**All features except for "flower pixels"**	**All features except for estimated number of flowers**
Linear Regression	91% (±0.03)	90% (±0.04)	85% (±0.05)
Huber Regression	91% (±0.02)	90% (±0.02)	85% (±0.06)

Total number of observations, number of observations in training data, and number of observations in testing data are given. 5-fold cross validation was used. Input features for inflorescence branch number prediction were plant age, "flower pixels," and number of flower components (estimated number of flowers). SD = Standard deviation across cross validation replicates.

As some input features were correlated [flower pixels and number of flower components (estimated number of flowers), Spearman's correlation coefficient: $r^2 = 0.99$], the prediction was tested without the flower pixels feature and achieved 90% ± 0.03 accuracy. Overall, the number of the flower components feature seemed to have more impact on branch prediction accuracy than the flower pixels feature. Consecutively, individual components from the thresholded blue-yellow channel were assigned to individual raceme branches (**Figure 4D**), by using hierarchical clustering. Taken together with additional image-derived traits such as canopy width and plant height, this allowed canopy architecture for the different genotypes included in this experiment to be broadly characterized (**Supplementary Figure 3.4**).

Drought Phenotyping

It was anticipated that withholding water from a subset of actively growing plants would result in a reduced growth rate, an increase in temperature resulting from reduced evapotranspiration, and wilting from reduced turgor pressure. In order to test semi-automated detection of these symptoms, drought stress was visually scored for all plants of both control and drought-treatment groups using a binary scheme (0: unstressed, 1: symptoms of drought stress) and the top and side view RGB images on DAS 49.

First, three individual image-based traits were tested for their efficacy as proxy traits in semi-automated drought stress detection: NIR intensity, Excess Green Index (ExG), and Δ pixelsTV (the change in the number of plant pixels from top view between DAS 47 and DAS 49, assuming that wilting would result in a smaller number of plant pixels captured from the top). However, taken on their own, none of these three traits showed substantial differences in plants displaying visible drought stress compared to unstressed plants on DAS 49 (**Supplementary Figure 3.6**). Similarly, two-way combinations of any of the three traits could not reliably distinguish between stressed and non-stressed plants (data not shown).

Next, curvature of the segmentation boundary was tested as an indicator of drought stress. The histogram of curvature over scale (HOCS; Kumar et al., 2012) method was applied to the top view plant boundary on DAS 49, assuming that plants experiencing water deficit would exhibit rolling of leaf edges. Although the histograms of curvature of stressed and unstressed plants did exhibit differences in their proportions at different scales, their within-group variation was too high to reliably distinguish between these two phenotypes (**Figure 9**).

Finally, the potential of machine learning algorithms in semi-automated drought stress detection was assessed by comparing the results of supervised machine learning to the manually determined stress scores of DAS 49. Using only the curvature of the top view plant boundary (summarized by HOCS), the highest accuracy achieved in initial tests was 66.6% using the KNN algorithm (K = 1), which was judged too low to be investigated any further. In contrast, using a combination of 21 phenotypic traits and image attributes extracted from the image dataset (mean and 75th percentile NIR, total number of plant pixelsTV, Δ pixelsTV, the convex hull areaTV, Δ convex hullTV, total number of plant pixelsSV, Δ pixelsSV, the convex hull areaSV, Δ convex hullSV, weight before, and after watering, plant area below pot rim, plant height, plant width, Δ plant height (between DAS 47 and 49), Δ plant width, ExG, flowering group (early, intermediate, late), and number of raceme branches) with several supervised machine learning algorithms was able to detect drought stress with the area under the receiver operating characteristic curve (ROC AUC) > 0.8 with 5-fold cross validation (**Table 2**). The maximum mean ROC AUC was achieved using the Random Forest algorithm (0.85 ± 0.03), Linear Discriminant Analysis (0.82 ± 0.04), and Logistic Regression (0.82 ± 0.08) algorithms. We also calculated the accuracy and achieved a maximum mean accuracy of 81% for 5-fold cross validation with the Random Forest Algorithm. Removing single features resulted only in very small changes to overall identification of ROC AUC in most cases (**Supplementary Table 3.4**).

DISCUSSION

The potential of HTP combined with automated image processing and machine learning for accelerating crop improvement has been increasingly acknowledged; however, deployment is currently hampered by the lack of reliable, robust, and easy to implement data processing methods. The results presented here show that scoring of basic plant traits, such as plant height and width, but also more complex traits such as flowering traits and canopy architecture in a streamlined, semi-automated fashion can be reliably achieved, once a workflow has been optimized. In addition, the potential of machine learning to identify multi-dimensional plant phenotypes such as drought stress resistance was shown. This methodology thus shows great promise for application in targeted breeding programs for *B. napus* and other crops.

Introduction of the Diverse Founder Panel for SKBnNAM

Since their introduction more than a decade ago, NAM populations have proven to be invaluable tools for the dissection of the genetic architecture of complex traits and have been adopted for a number of economically important crops (Yu et al., 2008; Gage et al., 2020). The lines phenotyped in this study form the founder panel for a new spring-type *B. napus* NAM population, SKBnNAM. Besides the common founder, N99-508, this panel includes seven Canadian adapted, five Australian, 13 European, 18 Asian as well as seven other (e.g., Argentinian and yellow-seeded) lines (**Supplementary Table 3.2**), which capture the genetic spectrum of spring-type *B. napus* (**Figures 5**, **6**). This population represents a useful complement to the previously published Bn-NAM population (16 founder lines, 2425 F_6 recombinant inbred lines) which was developed from mostly semi-winter-type rapeseed, a morphotype that is predominantly grown in China (Hu et al., 2018), as well as to a large *B. napus* NAM population developed from winter and synthetic lines (51 founder lines, Snowden et al., 2015). Based on preliminary genotype data and available field

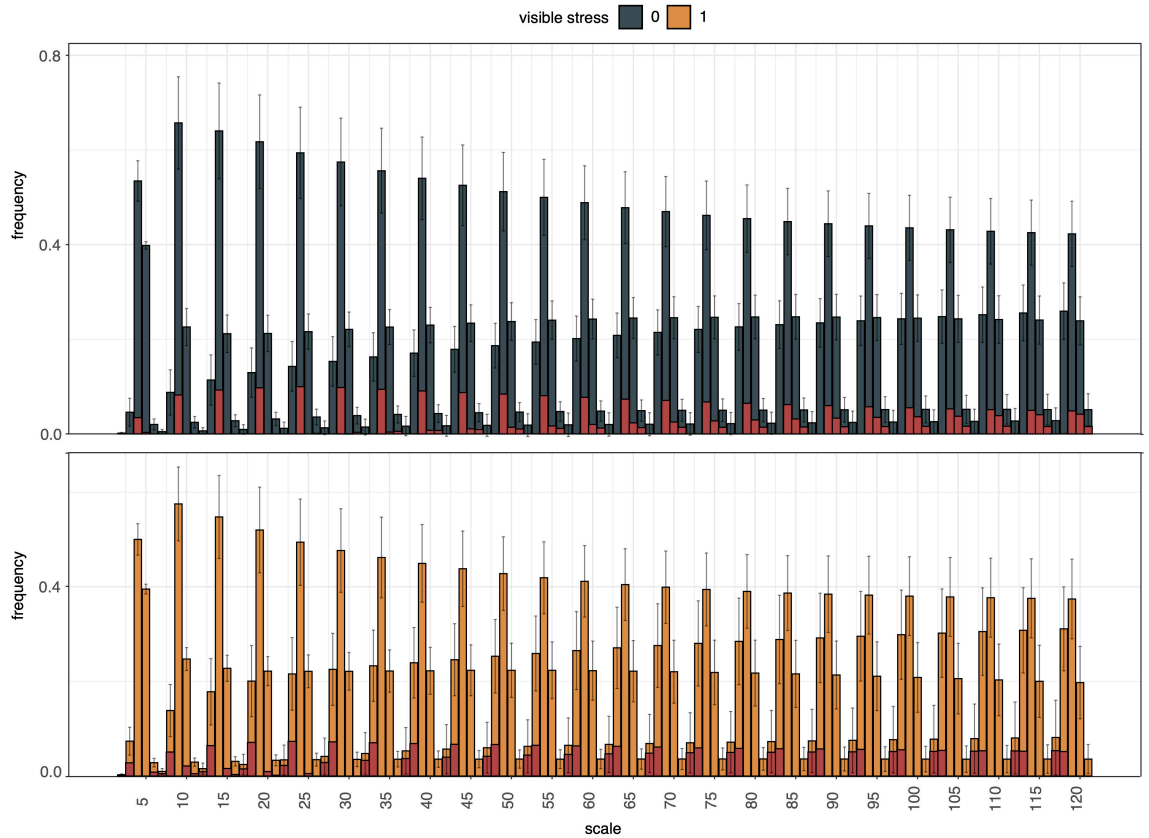


FIGURE 9 | Summary of histogram of curvature over scale (HOCS) method for all unstressed (0, blue) and all visibly stressed (1, orange) individuals on DAS 49. Each plot shows a series of 25 histograms that are composed of 5 bins each summarizing the curvature of the segmentation boundary at a given scale. Error bars give standard deviations across all individuals in each group (unstressed, visibly stressed). Overall histogram distributions were visibly different between these two groups (highlighted by red column portions signifying absolute mean differences for each bin). However, high within-group variation across the diversity panel masked these between-group differences and prevented accurate drought stress prediction from curvature of segmentation boundaries alone.

TABLE 2 | Mean ROC AUC, and mean accuracies for drought stress identification from image-derived traits for 50 *B. napus* genotypes using different machine learning algorithms.

	Drought stress identification	
Number of down-sampled observations (total/training data/testing data)	266/212/54	
	Mean ROC AUC with 5-fold cross validation	**Mean accuracy with 5-fold cross validation (SD)**
Random Forest	**0.85**	**0.81 (±0.03)**
Linear Discriminant Analysis	0.82	0.77 (±0.04)
Logistic Regression	0.82	0.79 (±0.08)
Decision Tree	0.73	0.75 (±0.12)
KNN classifier (K = 1)	0.72	0.73 (±0.08)
Support Vector Machine (SVM)	0.76	0.67 (±0.23)

Total number of observations, number of observations in training data, and number of observations in testing data are given. 5-fold cross validation was used. Input features for drought stress identification on 49 DAS were mean and 75th percentile NIR, total number of plant pixelsTV, Δ pixelsTV, the convex hull areaTV, Δ convex hullTV, total number of plant pixelsSV, Δ pixelsSV, the convex hull areaSV, Δ convex hullSV, weight before, and after watering, plant area below pot rim, plant height, plant width, Δ plant height, Δ plant width, ExG, flowering group (early, intermediate, late), and number of raceme branches. All Δ plant values represented differences between 47 DAS and 49 DAS. ROC AUC = Area Under the Receiver Operating Characteristic Curve. SD = Standard deviation across cross validation replicates. Achieved maximum results have been highlighted in bold.

phenotype data it was expected that SKBnNAM would be a valuable resource for dissecting the genetic architecture of complex traits in *B. napus*, and thus be instrumental for rapeseed crop improvement. However, it was still necessary to establish a clear picture of the level of phenotypic variation potentially segregating within the population.

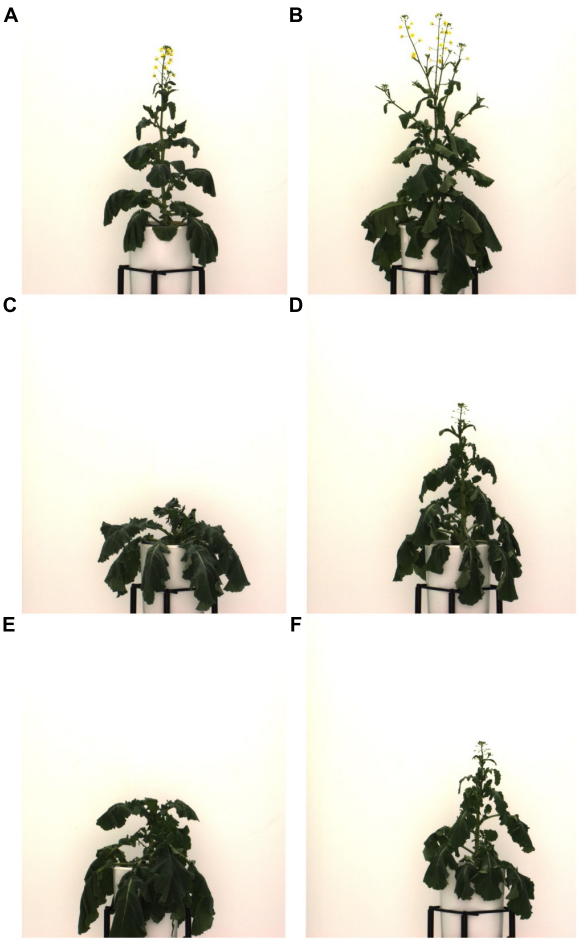


FIGURE 10 | Examples of *B. napus* individuals exhibiting mild to moderate drought stress on 47 DAS. All images show plants from control group. For further examples of drought stress symptoms, see **Supplementary Figure 3.3**. **(A)** NAM 12, Ind 423; **(B)** NAM 0, Ind 497; **(C)** NAM 30, Ind 492; **(D)** NAM 75, Ind 422; **(E)** NAM 33, Ind 479; and **(F)** NAM 75, Ind 457.

Considerations for Controlled Environment HTP Experiments in *B. napus*

The data presented represents one of the longest controlled environment HTP datasets for *B. napus* published to date. While this dataset will certainly become a valuable resource for image analysis refinements and innovations in rapeseed phenotyping, experimental problems that were encountered should be taken into consideration for future experiments. As mentioned above, higher than expected within genotype × treatment variation suggested that watering regimes were not administered as precisely as planned, leading, for example, to some individuals of the control group exhibiting symptoms of drought stress (**Supplementary Figures 3.2,3.3**). This was likely due to a combination of relatively small pot sizes (9 L pot volume, **Supplementary Table 3.3**) and automated overhead irrigation which led to larger rosette leaves effectively shielding the soil in some individuals. In addition, the automated plant-to-sensor movement within the phenotyping facility presented additional challenges during the experiment, as some plants were damaged (some larger, overhanging rosette leaves were torn, **Supplementary Figure 3.3**) while being moved through the system. Most of these challenges are connected to the plant architecture of *B. napus* as compared to other crop species frequently used in indoor HTP experiments (e.g., maize and rice) which suggests that scaling this system to different crop species requires crop-specific expertise and adjustments. Despite these challenges it was possible to focus on plant traits that were unaffected and further test the feasibility of different image analysis approaches for *B. napus*.

Flower Phenotyping/Growth Stages

Major breeding objectives in *B. napus* include early germination, improved seedling root development, early onset of flowering, high yield, favorable aerial plant architecture, reduced pod shatter, improved seed oil content, and quality as well as tolerance to biotic and abiotic stresses (Delourme et al., 2018). A focus of the current study was the semi-automated evaluation of flowering traits, plant aerial architecture and, where feasible, resistance to drought stress. It was demonstrated that it is possible to efficiently and accurately track flowering time and intensity through time using overhead RGB imaging (**Figure 8**). While we are not aware of any previous indoor HTP studies of *B. napus* that proposed flower detection and quantification to a similar degree of precision as this one, the approach presented here was similar to the one outlined by Chen et al. (2019) who separated flower volume from plant volume on the basis of pixel color. However, flowering traits were not the primary focus of their study and the authors did not investigate the possibility of using this method to track flower timing throughout their experiment. The results demonstrated that anthesis in *B. napus* does not need to be scored manually in greenhouse settings [as done by Chen et al. (2019)] but can be accurately detected semi-automatically (97.14% accuracy). This represents a significant extension of previous approaches as it further allows us to approximate flower numbers per plant and phenotyping time point. Accurate flower tracking also enabled the characterization of plant aerial architecture throughout the growing period (e.g., canopy width, canopy height, canopy angle, as well as number of raceme branches, **Figure 7**, and **Table 1**). Although not possible in the current experiment, the same methodology could easily be applied to capture the end of the flowering period and thus compare the length of the flowering period across genotypes and treatments. Semi-automated and eventually fully automated detection of flowering will facilitate high-throughput germplasm screening for early flowering varieties as well as extraction of other informative image traits or spectral indices that are scored during flowering, for example leaf spectral features which have been shown to be strongly correlated to seed yield (R^2 = 0.71) when measured during the flowering stage (Zhang and He, 2013). Due to a number of additional factors (e.g., in-field variability of growing conditions, presence of parasites/disease agents, spectral signal of canopies rather than single plants, and shadows from overlapping plants), field phenotyping requires a

slightly different approach to automated tracking of flowering in rapeseed. However, similar to our findings, the blue and green spectral bands of RGB images (e.g., used to calculate the normalized difference yellowness index, NDYI) have proven to be the most useful for distinguishing between flower and overall canola canopy signals in field settings (Sulik and Long, 2020).

Drought Phenotyping

Drought stress caused by moderate loss of water is characterized by a reduction in water content, diminished leaf water potential and turgor loss, accompanied by stomata closure, and decrease in cell enlargement and growth, while desiccation (severe water stress) can result in arrest of photosynthesis and even plant death (Jaleel et al., 2009). However, there is considerable plasticity in the degree to which water loss can be minimized and high-water status can be sustained among and within species (Farooq et al., 2013). This can lead to a broad spectrum of drought stress symptoms among closely related crop varieties. These wide-ranging symptoms are often visually scored by experienced breeders, but this integrated task is much harder to translate into automated stress detection protocols.

Using single or dual image-based plant traits such as NIR intensities, ExG (a measure of greenness), and daily changes in projected leaf area (Δ pixelsTV) were not reliable ways to identify drought stress in our experiment (**Figure 6** and **Supplementary Figure 3.6**). This is in contrast to findings from other indoor phenotyping experiments. For example, Janni et al. (2019) found significantly different NIR intensities and greenness values for drought-stressed and well-watered tomato plants. Similarly, Vello et al. (2015) found that they could accurately identify *A. thaliana* individuals in a water-limited treatment group using an NIR intensity threshold. Finally, Briglia et al. (2019) reported a clear correlation between stem water potential and both, NIR intensities and greenness parameters for drought-treated grapevines. One likely explanation for this difference is the degree to which drought was imposed during these experiments. While plants were often stressed irreversibly, to the extent of wilting, leaf yellowing or browning (senescence), and even lethal dehydration in previous studies (e.g., Vello et al., 2015; Duan et al., 2018; Briglia et al., 2019), individuals exhibited only comparatively mild drought symptoms in our experiment (e.g., leaf rolling, loss of turgor, **Figure 10**, and **Supplementary Figure 3.3**). It thus appears that both NIR intensities and greenness indices such as ExG are more suitable for detecting severe levels of drought stress. In addition, the second half of the treatment phase coincided with the onset of flowering leading to temporal overlap in flowering and more severe drought stress. This created substantial problems for the utility of NIR for drought detection since yellow *B. napus* flowers refract NIR the same way that senescent leaves would and thus, flowering-dependent changes in NIR intensities effectively mask the effects of drought stress (also seen by a significant positive relationship between NIR intensity and the number of flower pixels, Spearman's rank correlation rho = 0.38, **Supplementary Figure 3.7A**). Unsurprisingly, ExG was also driven by presence of flowers (rho = −0.68, **Supplementary Figure 3.7B**). Thus, the suitability of previously applied image-based traits for measuring drought stress strongly depends on experimental timing of drought stress and flowering in *B. napus*. Finally, very few previous drought phenotyping studies included diversity panels of the studied crops. The genotypic variation in leaf color, leaf thickness, leaf shape and size, and plant size made it impossible to define common thresholds of the tested parameters that would have allowed accurate semi-automated drought stress identification across all 50 assayed *B. napus* lines (**Supplementary Figure 3.8**). This illustrates the limitations of using these traits in complex experimental settings, such as diverse pre-breeding trials.

Instead, the data revealed the promise of supervised machine learning for stress identification in such complex situations. To the best of our knowledge, plant boundary curvature as summarized by the HOCS method, was used here for the first time to evaluate drought stress in an indoor HTP experiment. While promising overall, using this metric on its own was judged to be not effective for reliably distinguishing between stressed and unstressed plants, again likely due to genotype-specific variation overriding any shared symptoms of drought stress. However, using combinations of plant traits and image attributes derived from RGB and NIR images, it was possible to achieve stress identification accuracies of more than 80%. This mirrors previous studies that reported promising results using a variety of machine learning algorithms for drought detection and prediction in other crops (e.g., Römer et al., 2012; Raza et al., 2014); however, here for the first time with the added complexity of a large diversity panel. This added complexity clearly identified drought stress to be a multi-dimensional phenotype that could not be described by single attributes or traits but instead required a combination thereof. Compared to hyperspectral imaging, which has successfully been used in drought stress identification (e.g., Römer et al., 2012; Susič et al., 2018; Asaari et al., 2019), but which can be substantially more costly and difficult to acquire, process, and store (Holzapfel, 2007), our results underline the utility of NIR and multispectral RGB for this kind of problem.

Despite these promising results, several drawbacks remain to be solved. Image data preprocessing, one of the most crucial steps for successful use of machine learning methods (Singh et al., 2016), is time-consuming and requires deep-rooted knowledge of the organism of interest for input feature selection. For example, even though several different combinations of input features were tested (**Supplementary Table 3.4**), it was not possible to achieve stress identification ROC AUC of >85% in this experiment. Yet, further refinement of the input feature combinations for reliable drought stress identification, for example using additional plant attributes or traits could lead to higher identification accuracies. In addition, further improvements could likely be achieved through expansion of the manually scored stress ground truth to multiple days in order to allow for better genotypic-specific trait learning. However, this would require a significant time investment as manually scoring mild drought symptoms from images can be relatively tedious and somewhat subjective even to the experienced eye. The ultimate goal of stress phenotyping is stress prediction before the onset of stress symptoms that can be distinguished by the human eye. Even though our study makes a valuable contribution to this overall problem, future work will be needed to solve the puzzle of drought stress prediction in *B. napus* and other crops.

CONCLUSION

Phenotyping is considered a major bottleneck slowing the development of new crop varieties. This study illustrates the utility of semi-automated image processing and supervised machine learning for pre-breeding activities in *B. napus* by demonstrating their efficacy in scoring key agronomic traits including flowering characteristics (e.g., timing and volume), canopy architecture traits (e.g., raceme branch numbers), and early symptoms of drought stress in a diverse panel of spring lines. Despite several methodological challenges connected to scaling an indoor HTP platform to different crops, the results presented underline the promise of state-of-the-art HTP technologies, semi-automated image processing, and supervised machine learning for crop improvement, in particular when combined with genomics and systematic breeding strategies.

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://p2irc-data-dev.usask.ca/dataset/10.1109.SciDataManager.2020.7284788, N/A; https://genomevis.usask.ca/haplotype-map-tree/, N/A.

AUTHOR CONTRIBUTIONS

IP and SR conceived and supervised the original experiments and research plans. NK supervised by IM designed and performed image processing, trait extraction, and machine learning experiments. JE carried out data analysis, generated all data annotations, and wrote article with contributions from all the authors. VB and CG conceived and built the SNP data visualization tool. SV was involved in NAM founder line selection and original phenotyping. EH contributed SNP genotyping and genomic mapping data and provided technical assistance for their analysis. KH assisted with founder line selection and further population development. IP agreed to serve as the author responsible for communication. All authors contributed to the article and approved the submitted version.

FUNDING

The SKBnNAM population was developed with funding from the Saskatchewan Agricultural Development Fund, SaskCanola, Alberta Canola Producers, and three commercial entities. In addition, we would like to acknowledge the funding for the phenotyping work: Canada First Research Excellence Fund through the "Designing Crops for Global Food Security" at the University of Saskatchewan.

ACKNOWLEDGMENTS

We would like to thank the AAFC canola breeding field team for assistance with field phenotyping, in particular Brad Hope, Murray Lewis, Paul Prodahl and Ryan Vetter.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.780250/full#supplementary-material

REFERENCES

Asaari, M. S. M., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., and Scheunder, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* 162, 749–758.

Becker, H. C., Engqvist, G. M., and Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor. Appl. Genet.* 91, 62–67. doi: 10.1007/BF00220859

Berger, B., Parent, B., and Tester, M. (2010). High-throughput shoot imaging to study drought responses. *J. Exp. Bot.* 61, 3519–3528. doi: 10.1093/jxb/erq201

Briglia, N., Montanaro, G., Petrozza, A., Summerer, S., Cellini, F., and Nuzzo, V. (2019). Drought phenotyping in Vitis vinifera using RGB and NIR imaging. *Sci. Hortic.* 256:108555.

Chen, S., Guo, Y., Sirault, X., Stefanova, K., Saradadevi, R., Turner, N. C., et al. (2019). Nondestructive phenomic tools for the prediction of heat and drought tolerance at anthesis in *Brassica* species. *Plant Phenomics* 2019, 1–16. doi: 10.34133/2019/3264872

Choudhury, S. D., Bashyam, S., Qiu, Y., Samal, A., and Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* 14:35. doi: 10.1186/s13007-018-0303-x

Choudhury, S. D., Stoerger, V., Samal, A., Schnable, J. C., Liang, Z., and Yu, J.-G. (2016). "Automated vegetative stage phenotyping analysis of maize plants using visible light images," in *KDD workshop on Data Science for Food, Energy and Water (KDD -DSFEW)*, (San Francisco, CA). doi: 10.1186/s13007-018-0303-x

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., et al. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899. doi: 10.1007/s00122-016-2746-7

Cruz, J. A., Yin, X., Liu, X., Imran, S. M., Morris, D. D., Kramer, D. M., et al. (2016). Multi-modality imagery database for plant phenotyping. *Mach. Vis. Appl.* 27, 735–749.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Delourme, R., Laperche, A., Bouchet, A.-S., Jubault, M., Paillard, S., Manzanares-Dauleux, M.-J., et al. (2018). "Genes and Quantitative Trait Loci Mapping for Major Agronomic Traits in *Brassica napus* L," in *The Brassica napus* Genome, eds S. Liu, R. Snowdon, and B. Chalhoub (Cham: Springer), 41–85.

Diers, B. W., and Osborn, T. C. (1994). Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theor. Appl. Genet.* 88, 662–668. doi: 10.1007/BF01253968

Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., et al. (2018). Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Front. Plant Sci.* 9:492. doi: 10.3389/fpls.2018.00492

Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., et al. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in *Setaria*. *Mol. Plant* 8, 1520–1535. doi: 10.1016/j.molp.2015.06.005

Farooq, M., Hussain, M., Wahid, A., and Siddique, K. H. M. (2013). "Drought Stress in Plants: an Overview,"in *Plant Responses to Drought Stress: from Morphological to Molecular Features*, ed. R. Aroca (Berlin Heidelberg: Springer-Verlag).

Freedman, D. A. (2009). *Statistical Models: theory And Practice*. Cambridge: Cambridge university press.

Friedt, W., Tu, J., and Fu, T. (2018). "Academic and economic importance of *Brassica napus* rapeseed," in *The Brassica napus* genome, eds S. Liu, R. Snowdon, and B. Chalhoub (Berlin: Springer), 1–20.

Gage, J. L., Monier, B., Giri, A., and Buckler, E. S. (2020). Ten years of the Maize Nested Association Mapping population: impact, limitations, and future directions. *Plant Cell* 32, 2083–2093. doi: 10.1105/tpc.19.00951

Gazave, E., Tassone, E. E., Ilut, D. C., Wingerson, M., Datema, E., Witsenboer, H. M. A., et al. (2016). Population genomic analysis reveals differential evolutionary histories and patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. *Front. Plant Sci.* 7:525. doi: 10.3389/fpls.2016.00525

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. (2017). PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088. doi: 10.7717/peerj.4088

Hickey, L. T., Hafeez, A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9

Holzapfel, C. B. (2007). *Estimating nitrogen fertilizer requirements of Canola (Brassica napus L.) using sensor-based estimates of yield potential and crop response to nitrogen*. Ph.D. thesis. Winnipeg: Department of Soil Science, University of Manitoba

Hu, J., Guo, C., Wang, B., Ye, J., Liu, M., Wu, Z., et al. (2018). Genetic properties of a nested association mapping population constructed with semi-winter and spring oilseed rapes. *Front. Plant Sci.* 9:1740. doi: 10.3389/fpls.2018.01740

Huber, P. J. (1992). "Robust estimation of a location parameter," in *Breakthroughs in statistics, Springer Series in Statistics (Perspectives in Statistics)*, eds S. Kotz and N. L. Johnson (New York, NY: Springer).

Jaleel, C. A., Manivannan, P., Wahid, A., Farooq, M., Somasundaram, R., and Panneerselvam, R. (2009). Drought stress in plants: a review on morphological characteristics and pigments composition. *Int. J. Agric. Biol.* 11, 100–105.

Janni, M., Coppede, N., Bettelli, M., Briglia, N., Petrozza, A., Summerer, S., et al. (2019). In vivo phenotyping for the early detection of drought stress in tomato. *Plant Phenomics* 2019, 1–10. doi: 10.34133/2019/6168209

Junker, A., Muraya, M. M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A. E., et al. (2015). Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant Sci.* 5:770. doi: 10.3389/fpls.2014.00770

Kjaer, K., and Ottosen, C.-O. (2015). 3D Laser triangulation for plant phenotyping in challenging environments. *Sensors* 15, 13533–13547. doi: 10.3390/s150613533

Knoch, D., Abbadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., et al. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnol. J.* 18, 68–82. doi: 10.1111/pbi.13171

Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., et al. (2012). "Leafsnap: a Computer Vision System for Automatic Plant Species Identification," in *Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7573*, eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Berlin, Heidelberg: Springer), 502–516.

Lenaerts, B., Collard, B. C. Y., and Demont, M. (2019). Improving global food security through accelerated plant breeding. *Plant Sci.* 287:110207. doi: 10.1016/j.plantsci.2019.110207

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

Manay, S., Cremers, D., Hong, B.-W., Yezzi, A. J., and Soatto, S. (2006). Integral invariants for shape matching. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 1602–1618.

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Shawarma, R., Schumann, E., et al. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* 16:290. doi: 10.1186/s12864-015-1459-7

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognit. Lett.* 81, 80–89.

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007

Pommerrenig, B., Junker, A., Abreu, I., Bieber, A., Fuge, J., Willner, E., et al. (2018). Identification of rapeseed (*Brassica napus*) cultivars with a high tolerance to Boron-deficient conditions. *Front. Plant Sci.* 9:1142. doi: 10.3389/fpls.2018.01142

R Core Team (2020). *R: a Language And Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing

Rahman, H. (2013). Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Can. J. Plant Sci.* 93, 363–373.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:66428. doi: 10.1371/journal.pone.0066428

Raza, S.-A., Smith, H. K., Clarkson, G. J. J., Taylor, G., Thompson, A. J., Clarkson, J., et al. (2014). Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery. *PLoS One* 9:e97612. doi: 10.1371/journal.pone.0097612

Rebetzke, G. J., Jimenez-Berni, J., Fischer, R. A., Deery, D. M., and Smith, D. J. (2019). Review: high-throughput phenotyping to enhance the use of crop genetic resources. *Plant Sci.* 282, 40–48. doi: 10.1016/j.plantsci.2018.06.017

Römer, C., Wahabzada, B. M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., et al. (2012). Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. *Funct. Plant Biol.* 39, 878–890. doi: 10.1071/FP12060

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015

Snowden, R. J., Abbadi, A., Kox, T., Schmutzer, T., and Leckband, G. (2015). Heterotypic Haplotype Capture: precision breeding for hybrid performance. *Trends Plant Sci.* 20, 410–413. doi: 10.1016/j.tplants.2015.04.013

Song, Q., Yan, L., Quigley, C., Jordan, B. D., Ficus, E., Schroeder, S., et al. (2017). Genetic characterization of the soybean Nested Association Mapping population. *Plant Genome* 10, 1–14. doi: 10.3835/plantgenome2016.10.0109

Sulik, J. J., and Long, D. S. (2020). Automated detection of phenological transitions for yellow flowering plants such as *Brassica* oilseeds. *Agrosyst. Geosci. Environ.* 3:e20125. doi: 10.1002/agg2.20125

Susič, N., Žibrat, U., Širca, S., Strajnar, P., Razinger, J., Knapič, M., et al. (2018). Discrimination between abiotic and biotic drought stress in tomatoes using hyperspectral imaging. *Sens. Actuators B Chem.* 273, 842–852. doi: 10.1016/j.mex.2019.02.022

Thomas, C. L., Alcock, T. D., Graham, N. S., Hayden, R., Matterson, S., Wilson, L., et al. (2016a). Root morphology and seed and leaf ionomic traits in a *Brassica napus* L. diversity panel show wide phenotypic variation and are characteristic of crop habit. *BMC Plant Biol.* 16:214. doi: 10.1186/s12870-016-0902-5

Thomas, C. L., Graham, N. S., Hayden, R., Meacham, M. C., Neugebauer, K., Nightingale, M., et al. (2016b). High-throughput phenotyping (HTP) identifies seedling root traits linked to variation in seed yield and nutrient capture in field-grown oilseed rape (*Brassica napus* L.). *Ann. Bot.* 118, 655–665. doi: 10.1093/aob/mcw046

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

Van Vliet, L. J., and Verbeek, P. W. (1993). "Curvature and bending energy in digitized 2D and 3D images," in *Proceedings of the 8th Scandinavian Conference on Image Analysis*, eds K. A. Høgda, B. Braathen, and K. Heia (Tromso: NOBIM), 1403–1410.

Veley, K. M., Berry, J. C., Fentress, S. J., Schachtman, D. P., Baxter, I., and Bart, R. (2017). High-throughput profiling and analysis of plant responses over time to abiotic stress. *Plant Direct* 1:e00023. doi: 10.1002/pld3.23

Vello, E., Tomita, A., Diallo, A. O., and Bureau, T. E. (2015). A comprehensive approach to assess Arabidopsis survival phenotype in water-limited condition

using a non-invasive high-throughput phenomics platform. *Front. Plant Sci.* 6:1101. doi: 10.3389/fpls.2015.01101

Wickham, H. (2016). *ggplot2: elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Woebbecke, D. M., Meyer, G. E., von Bargen, K., and Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* 38, 259–269.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008

Yu, J., Holland, J. B., Mcmullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in Maize. *Genetics* 178, 539–551. doi: 10.1534/genetics.107.074245

Zhang, X., and He, Y. (2013). Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves. *Ind. Crops Prod.* 42, 416–420.

Zhang, Y., Thomas, C. L., Xiang, J., Long, Y., Wang, X., Zou, J., et al. (2016). QTL meta-analysis of root traits in *Brassica napus* under contrasting phosphorus supply in two growth systems. *Sci. Rep.* 6:33113. doi: 10.1038/srep33113

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: current status and perspectives. *Front. Plant Sci.* 10:714. doi: 10.3389/fpls.2019.00714

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., and Weir, B. (2012). A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

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 in Plant Science

ORIGINAL RESEARCH
published: 15 February 2022
doi: 10.3389/fpls.2021.798371

Heterosis Derived From Nonadditive Effects of the *BnFLC* Homologs Coordinates Early Flowering and High Yield in Rapeseed (*Brassica napus* L.)

Caochuang Fang[1], Zhaoyang Wang[1], Pengfei Wang[1], Yixian Song[1], Ali Ahmad[1], Faming Dong[1], Dengfeng Hong[1,2] and Guangsheng Yang[1,2]**

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

Early flowering facilitates crops to adapt multiple cropping systems or growing regions with a short frost-free season; however, it usually brings an obvious yield loss. In this study, we identified that the three genes, namely, *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*, are the major determinants for the flowering time (FT) variation of two elite rapeseed (*Brassica napus* L.) accessions, i.e., 616A and R11. The early-flowering alleles (i.e., *Bnflc.a2* and *Bnflc.c2*) and late-flowering allele (i.e., *BnFLC.A3b*) from R11 were introgressed into the recipient parent 616A through a breeding strategy of marker-assisted backcross, giving rise to eight homozygous near-isogenic lines (NILs) associated with these three loci and 19 NIL hybrids produced by the mutual crossing of these NILs. Phenotypic investigations showed that NILs displayed significant variations in both FT and plant yield (PY). Notably, genetic analysis indicated that *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* have additive effects of 1.446, 1.365, and 1.361 g on PY, respectively, while their dominant effects reached 3.504, 2.991, and 3.284 g, respectively, indicating that the yield loss caused by early flowering can be successfully compensated by exploring the heterosis of FT genes in the hybrid NILs. Moreover, we further validated that the heterosis of FT genes in PY was also effective in non-NIL hybrids. The results demonstrate that the exploration of the potential heterosis underlying the FT genes can coordinate early flowering (maturation) and high yield in rapeseed (*B. napus* L.), providing an effective strategy for early flowering breeding in crops.

Keywords: heterosis, flowering time, *BnFLC*, non-additive effects, additive effects, rapeseed, plant yield

OPEN ACCESS

Edited by:
Ryo Fujimoto,
Kobe University, Japan

Reviewed by:
Javed Akhatar,
Punjab Agricultural University, India
Kun Lu,
Southwest University, China

****Correspondence:***
Dengfeng Hong
dfhong@mail.hzau.edu.cn
Guangsheng Yang
gsyang@mail.hzau.edu.cn

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

Received: *20 October 2021*
Accepted: *22 December 2021*
Published: *15 February 2022*

Citation:
Fang C, Wang Z, Wang P, Song Y, Ahmad A, Dong F, Hong D and Yang G (2022) Heterosis Derived From Nonadditive Effects of the BnFLC Homologs Coordinates Early Flowering and High Yield in Rapeseed (Brassica napus L.). Front. Plant Sci. 12:798371. doi: 10.3389/fpls.2021.798371

INTRODUCTION

Rapeseed (*Brassica napus* L.) is one of the most important oil crops worldwide and comprises more than 55% of the total oilseed production in China. But with the change of cultivation systems for next-stubble crops (e.g., rice and corn) in central China, the growth season of rapeseed is being restricted. And in northwestern China, undetermined early frost is often a threat to seed maturation. Thus, a shorter growing season or earlier maturation time would enable increased cultivation area and production for rapeseed in China. As flowering time (FT) is often associated with maturation time in crops (Wang Y. et al., 2021), the implementation of shorter growth stages *via* regulating FT becomes a feasible idea.

Advantageously, extensive genetic research on FT genes and qualitative trait loci (QTLs) has been conducted in rapeseed. In its close relative model plant *Arabidopsis*, nearly 306 genes have been reported in the FT regulatory network which are involved in vernalization pathway, autonomous pathway, photoperiod pathway, ambient temperature pathway, and gibberellins pathway (Bouché et al., 2016). This provides a reliable reference for FT in rapeseed (Schiessl et al., 2014; Bouché et al., 2016). In rapeseed, many QTLs and peaks identified *via* genome-wide association studies have been colocalized with some of the annotated FT genes from the reference genome (Schiessl, 2020), and the crucial genes, namely, *BnFLC.A10*, *BnFLC.A2*, and *BnFLC.C2*, which participate in the vernalization pathway, have been characterized in rapeseed (Chen L. et al., 2018; Yin et al., 2020).

However, many studies show that FT genes often induce pleiotropic effects in multiple traits including yield in some crops. In rice, *OsWOX13* not only triggers early initiation but also enhances drought tolerance (Minh-Thu et al., 2018). Mutations of *Osgi* and *Ospho2* delay initiation and reduce biomass but promote Pi concentration in rice (Li et al., 2017). The NAC transcription factor gene *OsY37* accelerates heading and also promotes leaf senescence in rice (Mannai et al., 2017). *OsWDR5a* interacts with *OsTrx1* to promote heading, secondary branches, and grain production (Jiang et al., 2018). *SDG701* promotes heading in rice, but some spikelets failed to produce seeds in heterozygous *sdg701-1 (+/–)* mutant plants, which decreased grain yield (Liu et al., 2017). *OsDHD1* interacting with *OsHAP5C/D* not only delays heading but also enhances yield (Zhang et al., 2019). *OsMFT1* delays heading in rice by suppressing *Ehd1*-, *FZP*-, and *SEPALLATA*-like genes, and increases the number of spikelets per panicle (Song et al., 2018). The *RFT 1* allele from rice cultivar Zhenshan 97 delayed heading and also increased plant height (PH), grain weight, number, and yield compared to Milyang 46 (Zhu et al., 2017). In soybean, the mutation *Glyma.04G050200.1*, which is an ortholog of the *Arabidopsis thaliana EARLY FLOWERING 3* gene, delayed flowering and also increased grain weight per plant under short-day conditions (Lu et al., 2017). In rapeseed, an association panel comprised of 195 inbred lines revealed that allelic variations of FT genes *FD* and *FES1* were associated with PH (Kaur et al., 2021). In the double haploid (DH) population derived from Skipton/Ag-Spectrum/Skipton rapeseed, FT was negatively correlated with yield but positively correlated with PH under long-day photoperiodic conditions (Raman et al., 2019). It is obvious that early FT alleles usually induced yield loss, and late-flowering alleles induced yield gain.

Interestingly, some studies showed that FT genes had a strong positive effect on the heterosis of yield traits. In tomatoes, *SFT* homologs to *AtFT* integrated the flowering signal and facilitated flowering. The hybrid between *SFT* and its mutant *sft* showed a strong heterosis which was originated from overdominance that increased yield by up to 60% (Krieger et al., 2010). In rapeseed, *BnFT* and *BnTFL1* paralogs not only promote flowering but also modulate yield-related productivity traits (Guo et al., 2014). Indeed, F_1 hybrids between $BnC6FTb_{G2154A}$ missense mutants and controls generated higher seed yields (Guo et al., 2014). Mutations of the *BnTFL1-2* paralog had no significant effects on FT but affected yield components. Indeed, F_1 hybrids between *BnTFL1-2* mutants and non-mutated parents had increased seed number per silique (SSN) and total seeds per plant, suggesting that heterozygous mutations in a *TFL1* paralog may impact heterosis in rapeseed (Guo et al., 2014).

Therefore, to adopt new cropping systems and short-season growing regions by rationally designing FT, the pleiotropic effect associated with FT genes, especially yield loss, should be carefully considered. Here, we dissected the major genetic components controlling the basis of FT variations between two elite inbreeding lines, i.e., 616A and R11, and produced a series of homozygous, as well as hybrids, to evaluate the genetic effects of the *FLC* homologs on FT and plant yield (PY). The results demonstrate that the exploration of heterosis from FT genes can coordinate early FT (early maturation) and high yield in rapeseed, providing an effective and indicative strategy for early maturation breeding in rapeseed and other crops.

MATERIALS AND METHODS

Plant Materials

Four *B. napus* inbred accessions were used in this study. 616A, the maternal line of several elite rapeseed hybrids in China, is an elite temperature-sensitive polima cytoplasmic male sterile (pol CMS) accession, which shows early flowering in both winter and spring environments. 616A is a DH line from the F_1 cross between a pol CMS line 987A (the female parent of a commercialized cultivar Heshengyou 868) and a pol CMS restorer 7,492 (Wang Z. et al., 2021). R11 acts as a semi-winter pol CMS restorer (Chen L. et al., 2018), which is an early-flowering mutation from the restorer R15 (the male parent of a commercialized hybrid cultivar Shengguang 101). ZS4R and 621R are also two semi-winter pol CMS restorers of several commercial hybrids released in China. A DH population with 352 independent genotypes is constructed from the F_1 between 616A and R11 by microspore culture. All materials were bred by Huazhong Agricultural University.

Field Experiment and Trait Evaluation

Field Experiments

The 352 DH lines and the two parents were randomly grown in two winter environments (i.e., Jiangling and Wuhan, Hubei) for two consecutive years, i.e., 2014–2015 and 2015–2016 seasons, and in one spring environment (Minle, Gansu) in the 2015 season. Two rows were set as a plot for each of the DH line, with 10 plants in each row.

The eight near-isogenic lines (NILs) (shown in the sections "Construction of Eight NILs for the FT Genes *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*" and "The Development of Eight NILs Differs at the Loci of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*"), 19 NIL hybrids (shown in the section "Heterozygous Status in *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* Generated Accumulative Heterosis in Plant Yield via Nonadditive "Effects"), and 16 non-NIL hybrids (shown in the section "Heterosis in Plant Yield Is Also Associated With Heterozygous *BnFLC.A2*, *BnFLC.C2*, and/or *BnFLC.A3b* Genotypes in

Non-NIL Hybrids") were planted as randomized blocks during the 2019–2020 season in two winter environments and one spring environment, with five plot repeats for the eight NILs and four plot repeats for the 19 NIL hybrids, 16 non-NIL hybrids, and two pol CMS restorers (i.e., ZS4R and 621R) in each of the three environments.

In all the field experiments, the row and plant spacing were 27 and 18 cm, respectively.

Phenotyping of Agronomic Traits of the Plots

FT was recorded as the number of days to flowering (DTF, in days) from sowing to the day when 50% of plants in the plot flowered. Eight plants for each plot of breeding tests in Jiangling and ten plants for each plot of breeding tests in Wuhan and Minle, excluding border rows, were randomly sampled to phenotype the following traits: branch number (BN), branch height (BH, distance from the cotyledonary node to the lowest branch in cm), plant height (PH, in cm), stem height (SH, distance from the cotyledonary node to the uppermost branch in cm), main raceme length (ML, in cm), silique number on main raceme (MSN), silique number per plant (PSN), PY (in g), silique length (SL, in mm), SSN, thousand seed weight (TSW, in g), oil content (OC, in %), protein content (PC, in %), and glucosinolate content (GC, in μmol/g).

DNA and RNA Preparation, Molecular Markers, and QTL Mapping

The DNA was extracted from leaves according to a modified CTAB method (Allen et al., 2006). RNA was extracted from the first leaf from the top of 616A and R11 in the 4-week-old plants cultivated in Wuhan using the Eastep^{TM} Super Total RNA Extraction Kit (Promega, Madison, WI, United States). cDNA was synthesized from RNA (4 μg samples) using the GoScript^{TM} Reverse Transcription Mix (Promega) and then analyzed using the GoTaq^{TM} qPCR Master Mix (Promega) and Bio-Rad CFX96 Real-Time System (Bio-Rad). The PCR program used was: 95°C for 1 min, and 40 cycles of 95°C for 10 s and 60°C for 30 s. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method based on three biological samples and three replicates for each sample (Livak and Schmittgen, 2001). The expression levels of *BnFLC.A3b* were monitored by the specific primer, as shown in **Supplementary Table 1** (Zou et al., 2012). The β-actin was used as the internal control (Wang et al., 2011).

The SCAR markers, i.e., STA2-55L/1R and STC2-4L/4R, were specifically amplified the wild-type (WT) alleles of *BnFLC.A2* and *BnFLC.C2*, respectively (Chen L. et al., 2018), while SCAR-5860/6430 was specifically amplified the nonfunctional allele of *Bnflc.a2* and *Bnflc.c2* in R11. *OPSNP7* was an intra-genetic marker specific for pol CMS restore gene *Rfp* (Liu et al., 2012). Insertion/deletion (InDel) markers for the candidate chromosomes were developed based on Illumina sequencing data of 616A and R11 referenced by the ZS11 genome (**Supplementary Table 1**; McKenna et al., 2010; Langmead and Salzberg, 2012; Song et al., 2020a). The 400 polymorphic SSR markers used for bulked segregant analysis (BSA) were obtained from a publication by Li et al. (2013). The sequence of the markers is shown in **Supplementary Table 1**.

The *BnFLC.A2*/*BnFLC.C2* and *Bnflc.a2*/*Bnflc.c2* subpopulations in the DH population from 616A and R11 were analyzed separately using the BSA strategy. For each of the subpopulation, the DNA of the 12 extremely early flowering and late-flowering individuals was used to build the bulk-DNA-pools, respectively. After the analysis by the 400 SSR markers, the co-segregated markers in the bulk-DNA pool, together with the InDel markers on chromosome A03, were used in the ICIM model and analyzed using QTL IciMapping software (Meng et al., 2015). The InDel markers on the candidate chromosomes were used to analyze the self-mapping population.

Construction of Eight NILs for the FT Genes *BnFLC.A2, BnFLC.C2,* and *BnFLC.A3b*

The marker SCAR-5860/6430 described in the section "DNA and RNA preparation, molecular markers, and QTL mapping" for the genetic analysis of 616A and R11 was used to trace *Bnflc.a2* and *Bnflc.c2*. The candidate gene was traced using the co-dominant InDel marker in the linkage map and tested using the SCAR marker developed according to the sequence variation of the candidate gene between 616A and R11. Otherwise, *OPSNP7* was used to trace *Rfp* and distinguish restore lines and sterile plants (Liu et al., 2012). The polymorphism SSR markers randomly selected from the BSA analysis in the section "DNA and RNA preparation, molecular markers, and QTL mapping" were used for background selection. The NILs of both fertile and sterile plants were self-pollinated continuously for at least two seasons, producing eight near-isogenic restorer lines and eight male sterile lines.

The MGISEQ-2000-high-throughput sequencing platform (PE150; MGI Tech Co. Ltd., Wuhan, China) was used to perform genome sequencing of the eight near-isogenic restore lines (Korostin et al., 2020). After the raw data were filtered by fastp (version 0.20.0) (Chen S. et al., 2018), the clean data were aligned to ZS11 reference genome to get the sequence alignment/map (SAM) format file by Bowtie 2 (version 2.4.1) (Langmead and Salzberg, 2012; Song et al., 2020a) and converted into binary alignment/map (BAM) format file using Samtools (version 1.9) according to the mapping coordinates (Li et al., 2009). The Picard tools (version 2.23.2) were used to remove the PCR duplicates and calculate the genome coverage[1]. By means of the Genome Analysis Toolkit (version 3.8), the single nucleotide polymorphisms (SNPs) that were not only homozygous in 616A and R11 but also heterozygous between 616A and R11 were determined as effective (McKenna et al., 2010). In Khan et al. (2018), the recurrent parent genome recovery analysis of the eight NILs, the SNP homozygous in 616A was defined as 0, the SNP homozygous in R11 was defined as 2, and the heterozygous SNP was defined as 1. A genomic region covered by 15 continuous SNPs with 1 SNP step length, set as a sliding window (Huang et al., 2009), was defined as a recovery region when they were continuously scored less than eight (the threshold of the breakpoint). In Khan et al. (2018), the recurrent parent genome

[1]http://broadinstitute.github.io/picard/

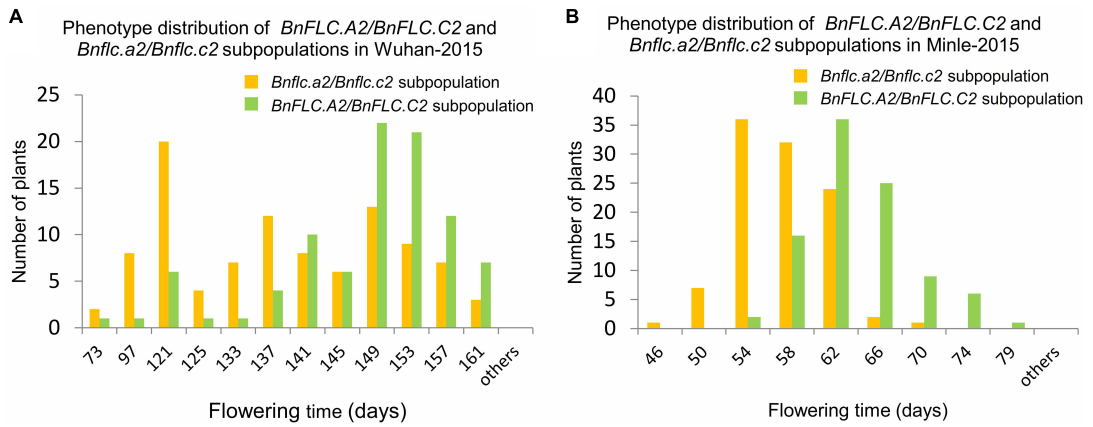


FIGURE 1 | Histogram of phenotype distribution of the subpopulations from the double haploid population constructed by crossing 616A and R11 in the Wuhan-2015 and Minle-2015 environments. **(A)** Phenotype distribution of *BnFLC.A2/BnFLC.C2* and *Bnflc.a2/Bnflc.c2* subpopulations in Wuhan-2015. **(B)** Phenotype distribution of *BnFLC.A2/BnFLC.C2* and *Bnflc.a2/Bnflc.c2* subpopulations in Minle-2015. The yellow color represented the *Bnflc.a2/Bnflc.c2* subpopulation, and the green color represented the *BnFLC.A2/BnFLC.C2* subpopulation.

recovery is defined by the length of the recovery region dividing the reference genome length of ZS11.

Statistical Analysis

The best linear unbiased estimation (BLUE) value was calculated using the restricted maximum likelihood (REML) linear mixed model in GenStat software. The genotypes of plant materials were set as fixed effects, whereas environment, plot, and randomly sampled plants were set as random effects.

The Pearson correlation coefficients were calculated by the BLUE values of the phenotypic data of the near-isogenic restorer lines and their hybrids using the package *PerformanceAnalytics* in R software (Dunne et al., 2018; Peterson and Carl, 2020).

Mid-parent heterosis was calculated *via* the BLUE values of hybrids and parental lines using the following formula: $\frac{H-\frac{1}{2}(p_1+p_2)}{\frac{1}{2}(p_1+p_2)} \cdot 100\%$ (Miedaner et al., 2016).

The genetic effects of the genes (QTLs) were calculated using the additive-dominant-epistasis model *via* the BLUE values of the phenotypic data of the NILs and their hybrids.

Multiple comparisons were conducted by least significance difference (LSD) using the base function in R.

RESULTS

Allelic Variations in Multiple *FLC* Paralogues Are Associated With FT Difference Between 616A and R11

Both R11 and 616A showed an early flowering phenotype, with the FT of 140–143 and 136–138 days in the winter environment, and 57–61 and 58–60 days in the spring environment, respectively. Though R11 harbors the mutated alleles (i.e., *Bnflc.a2* and *Bnflc.c2*) at both *BnFLC.A2* and *BnFLC.C2* loci due to the segmentally homeologous exchange (HE) between chromosomes A02 and C02 (Chen L. et al., 2018), it showed significantly later flowering than 616A in the winter environment and no significant difference of FT with 616A in the spring environment. At the same time, what is the genetic reason that 616A has the early flowering phenotype in both the winter and spring environments?

To understand the QTL (genes) conferring DTF variation between the parental accessions, we first analyzed the genotypes of *BnFLC.A2* and *BnFLC.C2* in 616A by STA2-55L/1R and STC2-4L/4R (**Supplementary Table 1**). The results showed that 616A carries the conserved WT alleles in both genes as reported before (Chen L. et al., 2018), which extremely significantly delayed the flowering in the DH population derived from the F_1 cross between 616A and R11 (**Supplementary Figures 1**, **2**). In addition, phenotypic variations in DTF remained wide in both of the WT "++/++" (*BnFLC.A2*/*BnFLC.C2*) and mutated " −−/−− " (*Bnflc.a2*/*Bnflc.a2*) subpopulations of the DH population (**Figure 1** and **Supplementary Figure 2**). These observations indicated that although typical sequence variations in *BnFLC.A2* and *BnFLC.C2* between 616A and R11 mediate a significant FT alteration, there must be some other genes controlling early flowering in 616A.

To explore the potential QTL (genes) rather than *BnFLC.A2* and *BnFLC.C2*, we constructed the extremely early and the late DNA bulks from the two subpopulations, respectively, and screened 400 polymorphic public SSR markers *via* the BSA strategy. In the two bulks, three co-segregated SSR markers, namely, SSR-1125, SSR-1873, and SSR-1979, were identified on chromosome A03 (**Supplementary Figure 1** and **Figure 2A**). Then, an InDel marker, InDel-A36M, was developed from this chromosome. By using these four markers, we constructed a local genetic map and identified an FT QTL named *qFT.A3* from the genomic region delimited by the markers, which explained 7.6297, 6.7333, and 5.2631% of the phenotypic variation in FT of the DH population in the Wuhan-2015, Wuhan-2016, and Minle-2015 environments, respectively (**Figure 2A**). The early FT allele was inherited from 616A.

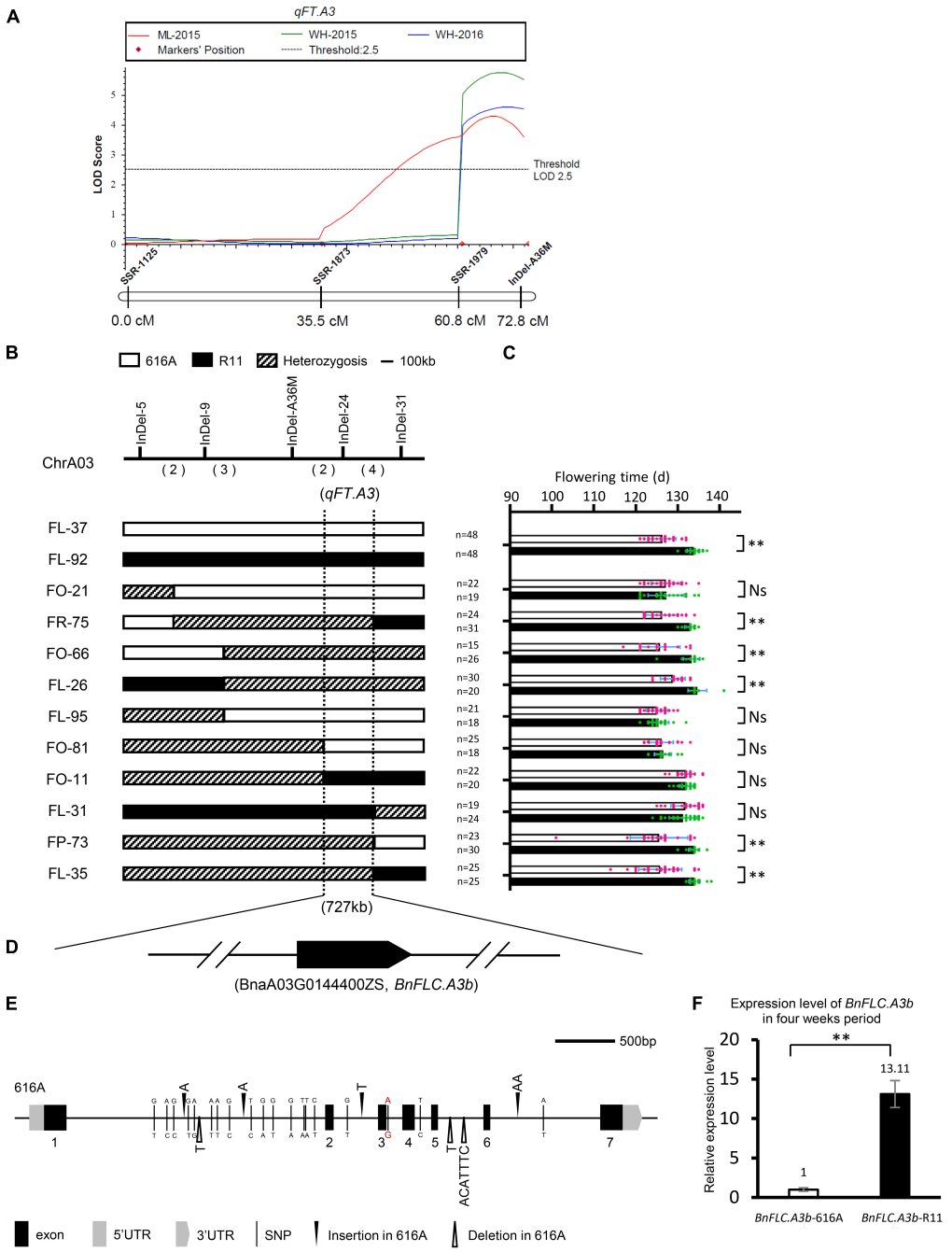


FIGURE 2 | Qualitative trait locus (QTL) mapping of *qFT.A3*. **(A)** Partial linkage map of chromosome A03 and QTL scanning of *qFT.A3*. **(B)** Genotype of the recombinants in the BC_2F_3 population. The numbers in parenthesis indicate the number of crossovers. **(C)** Progeny testing of two near-isogenic lines (NILs) and 10 recombinants narrowed *qFT.A3* to a region of 727 kb. Hollow bars indicate the phenotype of 616A homozygous segments and solid bars indicate the phenotype of R11 homozygous segments. Data represent mean ± SD. NS, not significant; **, significant at $p < 0.01$. **(D)** Annotated flowering time genes identified in the *qFT.A3* locus. Arrow indicated the direction of transcription; the break line indicated other genes not annotated for flowering time. **(E)** Sequence variations of the coding region of *BnFLC.A3b* in 616A and R11. The single nucleotide polymorphism and insertion/deletion variations between R11 and 616A were indicated by the benchmark of 616A. The red color indicated the G to A mutation of the five prime splicing sites between R11 and 616A. **(F)** Expression levels of *BnFLC.A3b* in 616A and R11, three biological replicates, and three technical replicates.

To narrow down the candidate interval of *qFT.A3*, more InDel markers were developed (**Supplementary Table 1**). Ten recombinant individuals were identified from the BC_2F_3 population between the InDel markers InDel-5 and InDel-31. The FT phenotypes of the corresponding BC_2F_4 progenies were investigated in the winter environment of Wuhan-2020. Based

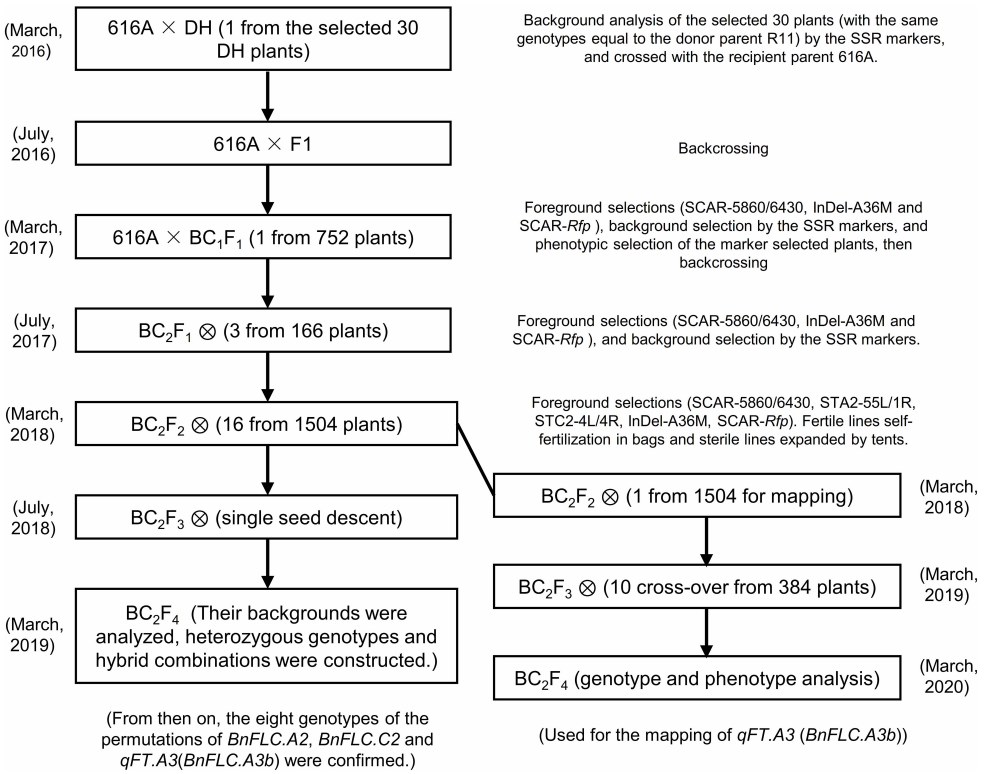


FIGURE 3 | Selection scheme for the construction of the eight genotypes.

TABLE 1 | Allele permutations, recurrent genome recoveries (Khan et al., 2018), and flowering time comparison of the eight genotypes.

Genotypes ID	Allele permutations			Genome coverage	Recovery rate	Multiple comparisons of flowering time		
	BnFLC.A2	*BnFLC.C2*	*qFT.A3 (BnFLC.A3b)*			Jiangling	Wuhan	Minle
ID01-000	0	0	0	79.90%	94.62%	65.6 a	58.0 a	50.2 a
ID02-010	0	1	0	78.60%	97.52%	111.4 b	89.8 b	53.4 b
ID03-001	0	0	1	80.68%	92.72%	113.0 b	90.4 b	54.2 b
ID04-100	1	0	0	79.76%	95.97%	123.8 c	102.2 c	59.4 c
ID05-011	0	1	1	78.92%	95.36%	133.0 d	128.6 d	59.8 c
ID06-110	1	1	0	77.84%	98.0%	140.8 e	137.2 e	63.8 d
ID07-101	1	0	1	74.14%	96.47%	143.6 e	138.4 e	66.6 e
ID08-111	1	1	1	76.56%	96.78%	148.0 f	140.6 f	68.6 f

0, mutated type allele; 1, wild-type allele.
The mutated alleles of BnFLC.A2 and BnFLC.C2 and the wild-type allele of qFT.A3 (BnFLC.A3b) were all from R11.
Letters on the right side of the numbers indicate the significance of multiple comparisons.

on this information, *qFT.A3* was further mapped into a 727-kb interval flanked by InDel-24 (**Figures 2B,C**). Genome annotation indicated that only one FT-related gene, *BnFLC.A3b*, resides in this candidate interval (**Figure 2D** and **Supplementary Table 2**; Song et al., 2020a).

Accordingly, we further amplified and comparatively analyzed the 1,691 bp promoter region and the coding region of *BnFLC.A3b* between the parental lines. Despite the conservation sequences in the seven annotated exons, seven InDels and 19 SNPs appeared in the introns, and five SNPs appeared in the promoter (**Figure 2E**). Among them, a G to A mutation in the five prime splicing site of the 3rd intron was very similar to the alternative splicing site of *FLC* orthologs, i.e., *BrFLC1* (6th intron G to A) and *BrFLC5* (3rd intron G to A), whose reduced expression levels were associated with the early FT of *Brassica rape* (Yuan et al., 2009; Xi et al., 2018). Consistently, we observed that the transcriptional level of *BnFLC.A3b* in R11 was nearly 13 times higher than that in 616A, suggesting the *BnFLC.A3b* allele in R11 was with a much stronger function than the allele in 616A (**Figure 2F**). In addition, similar sequence

difference and expressing variations of *BnFLC.A3b* were also reported between Ningyou7 and Tapidor (Zou et al., 2012). Based on these results, we speculated that besides *BnFLC.A2* and *BnFLC.C2*, *BnFLC.A3b* is most likely to be the third gene leading to the FT variation between 616A and R11. At last, a sequence-characterized amplified region (SCAR) marker named SCAR-A37M was subsequently developed according to the seven-base pairs InDel to specifically amplify the allele of *BnFLC.A3b*-R11 for the next genotyping of breeding materials (**Supplementary Table 1**).

The Development of Eight NILs Differs at the Loci of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*

According to the abovementioned analysis, it is clearly suggested that the allelic variations of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* at least partially determined the FT difference between 616A and R11. To carefully evaluate the sole and conjunctive effects of these genes, we attempted to develop NILs with the different allelic permutations in the 616A background by marker-assisted foreground and background selections (**Figure 3**). Considering that 616A is a pol CMS line, the nuclear restorer gene *Rfp* for pol-CMS was simultaneously introduced from R11 during backcross to facilitate the trait investigation in the later studies.

To reduce the breeding time, a DH line from the DH population coded as DH-8257, which shared the closest genetic background to 616A among the 30 DH lines with the same genotype as R11 at the three *FLC* target loci and *Rfp*, was chosen to backcross with 616A. After two rounds of foreground and background selection in BC_1F_1 and BC_2F_1 generations, the eight homozygous NILs of 616A and 616R (carrying the dominant *Rfp* allele) with permutations of *BnFLC.A2*, *BnFLC.C2*, and *qFT.A3* (*BnFLC.A3b*) were obtained in the resulting BC_2F_2 generation, respectively (**Figure 3** and **Table 1**). Then, a single seed descent (SSD) method was adopted to progress the generation to BC_2F_4. The analysis of the genome resequencing data showed that the genome recovery of these eight fertile NILs varied from 92.72 to 98.0% (**Table 1**). The eight fertile NILs with the 616R background were defined as the eight genotypes, which were numbered from ID01-000 to ID08-111, in which "0" and "1" individually represent the existence of the homozygous mutation and WT genotypes of any of the three *BnFLC* loci (**Figure 4** and **Table 1**).

Allelic Variations in *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* Had Additive Effects on FT as Well as in Seed Yield and Quality Trait in the Eight Genotypes

The FT of the eight genotypes was investigated in three environments, and significant differences were consistently observed in every condition (**Figure 4** and **Table 1**). ID01-000, possessing all the three mutated alleles at *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*, flowered earliest, whether in the winter environments (Jiangling and Wuhan) or in the spring environment (Minle) (**Table 1**), while ID08-111, carrying the WT genotypes at all loci, had the latest FT in all environments

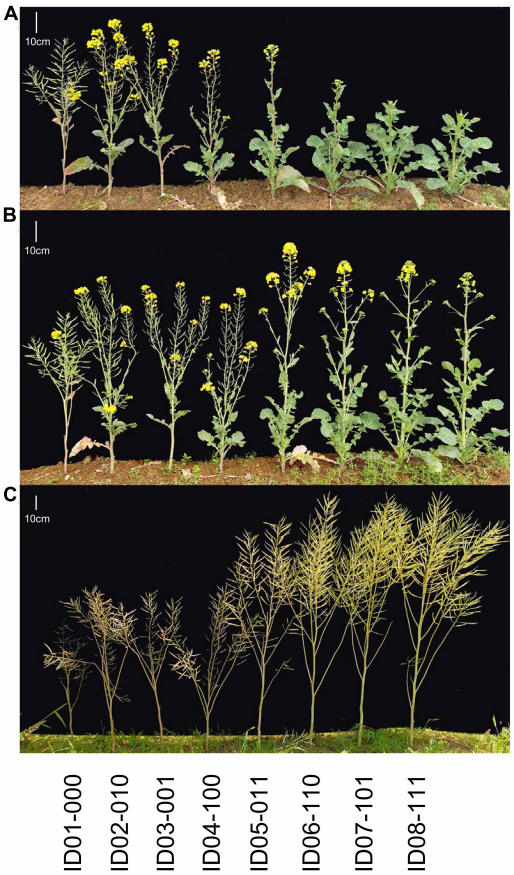


FIGURE 4 | Continuous observation of growth status for the eight genotypes in the Wuhan environment. **(A)** February 21, 2020. **(B)** February 29, 2020. **(C)** April 28, 2020. From left to right: ID01-000, ID02-010, ID03-001, ID04-100, ID05-011, ID06-110, ID07-101, and ID08-111.

(**Table 1**). Genotypes ID02-010, ID03-011, and ID04-100, which possessed only one WT allele, flowered later than genotype ID01-000, but earlier than other genotypes (**Table 1**). Genotypes ID05-011, ID06-110, and ID07-101, which possessed two of the WT alleles, flowered earlier than genotype ID08-111, but later than other genotypes (**Table 1**). The broad-sense heritability of FT in the Jiangling, Wuhan, and Mingle reached 0.8605, indicating that the eight genotypes exhibited a stable genetic effect on FT.

Compared to 616R (i.e., ID06-110), the other seven genotypes also exhibited widely phenotypic alterations in multiple traits (**Figure 4** and **Supplementary Figure 3**). Plant architecture traits, such as BN, BH, PH, and SH, showed a tendency from low to high with the delaying of FT (**Supplementary Figure 3**). Yield component traits, including MSN, PSN, SL, and SSN, were significantly lowered in early FT genotypes than in late FT genotypes (**Supplementary Figure 3**). Thus, ID01-000, with the earliest FT, had the lowest PY, while the latest FT ID08-111 had the highest PY, implying an obvious conflict between early FT and PY (**Supplementary Figure 3**). Similar changes were also found for OC, i.e., the genotypes showing earlier FT but sharing lower seed OC; in contrast,

Genetic effects	FT	BN	BH	PH	SH	ML	MSN	PSN	PY	SL	SSN	TSW	OC	PC	GC
m	95.74	7.206	42.45	144.8	85.36	59.44	56.73	257.8	20.0	61.65	22.14	4.629	44.44	21.79	35.38
a_{A2}	13.34	0.501	10.36	13.14	13.15	-0.010	5.311	14.52	1.446	0.263	0.532	-0.067	0.295	-0.548	-0.320
a_{C2}	8.571	0.376	4.0	7.40	7.972	-0.572	1.713	2.107	1.365	2.413	1.318	0.031	0.715	-0.722	0.135
a_{A3}	9.380	0.749	8.551	12.31	12.61	-0.298	4.645	12.53	1.361	1.968	1.258	-0.140	0.575	-0.566	-0.459
$a_{A2}a_{C2}$	-3.152	-0.092	0.094	-2.193	-0.844	-1.349	-0.868	1.068	-0.463	-0.572	-0.536	0.045	-0.761	0.655	0.288
$a_{A2}a_{A3}$	-2.846	0.067	-0.811	-3.494	-0.607	-2.887	-2.296	3.214	-0.281	-0.883	0.115	-0.122	-0.507	0.419	-1.361
$a_{C2}a_{A3}$	-2.554	0.078	-3.047	-2.607	-2.025	-0.583	-0.963	0.809	-0.407	1.157	-0.102	0.021	-0.652	0.445	-0.336
$a_{A2}a_{C2}a_{A3}$	-1.491	-0.022	-2.754	-2.250	-3.802	1.551	-1.286	-1.227	-0.282	-1.676	-0.729	-0.048	-0.217	-0.171	0.121
d_{A2}	0.647	0.410	2.132	6.471	5.296	1.129	3.823	37.30	3.504	0.156	0.026	-0.045	0.267	-0.069	0.220
d_{C2}	-1.263	0.363	1.103	4.092	3.775	0.271	1.968	34.29	2.991	-0.153	-0.922	6E-04	-0.464	0.639	-0.377
d_{A3}	-1.074	0.541	2.428	6.703	6.523	0.135	2.821	34.30	3.284	0.123	-0.260	-0.037	-0.304	0.333	-0.474
$a_{A2}d_{C2}$	0.025	0.198	1.069	2.794	3.123	-0.329	0.544	3.206	0.544	0.662	0.525	0.015	0.448	-0.335	0.894
$a_{A2}d_{A3}$	2.179	0.220	2.449	4.232	3.864	0.367	2.205	3.268	0.844	1.016	0.202	-0.004	0.442	-0.377	0.522
$d_{A2}a_{C2}$	0.426	0.169	2.699	1.635	3.621	-1.985	-0.369	-1.531	-0.264	-0.452	0.20	-0.149	-0.420	0.168	0.468
$a_{C2}d_{A3}$	-2.254	0.269	-0.896	0.053	0.358	-0.305	0.334	11.06	0.490	-0.97	0.260	-0.141	0.002	-0.111	0.088
$a_{A3}d_{A2}$	2.001	0.126	3.237	2.798	3.037	-0.239	0.543	-2.462	-0.169	-0.156	-0.046	0.012	0.288	-0.245	-0.220
$a_{A3}d_{C2}$	-0.414	-0.001	-0.263	-0.555	-0.602	0.048	-1.558	9.701	0.626	-1.066	0.222	-0.047	0.273	-0.239	-0.793
$d_{A2}d_{C2}$	1.413	-0.680	2.466	0.415	-0.948	1.408	-1.968	-26.07	-1.403	0.790	1.482	0.084	0.654	-0.734	-0.660
$d_{A2}d_{A3}$	-0.085	-0.769	-4.254	-8.014	-10.77	2.801	-2.660	-33.14	-3.071	0.649	-0.142	0.111	0.127	-0.120	0.127
$d_{C2}d_{A3}$	2.549	-0.655	0.838	-3.373	-4.592	1.265	-1.617	-30.96	-3.028	0.612	0.686	0.127	0.730	-0.708	0.126
$a_{A2}a_{C2}d_{A3}$	0.281	-0.142	-2.332	-0.899	-1.759	0.860	-0.206	-3.128	-0.414	0.138	0.275	-0.040	0.267	-0.245	-0.083
$a_{A2}d_{C2}a_{A3}$	1.108	0.172	1.465	2.226	2.846	-0.620	0.928	-3.90	0.098	0.134	0.464	-0.010	0.733	-0.622	0.949
$d_{A2}a_{C2}a_{A3}$	-1.203	-0.412	1.471	0.346	-0.818	1.164	-1.850	-6.474	-0.389	-0.788	0.264	-0.044	0.146	-0.356	1.033
$a_{A2}d_{C2}d_{A3}$	1.448	-0.146	0.553	-2.151	-0.717	-1.434	-0.662	-16.11	-1.657	-0.590	0.325	-0.087	-0.133	0.155	-1.037
$d_{A2}a_{C2}d_{A3}$	5.297	0.106	1.416	2.961	2.213	0.748	1.371	-7.560	0.909	2.320	0.188	0.289	0.518	-0.173	0.358
$d_{A2}d_{C2}a_{A3}$	3.863	-0.213	1.344	2.215	3.267	-1.051	3.107	-8.179	0.351	1.786	0.862	0.008	-0.690	0.861	0.865
$d_{A2}d_{C2}d_{A3}$	-2.114	1.008	-2.112	0.857	3.049	-2.237	3.711	55.21	4.393	-0.40	-1.393	-0.276	-1.042	0.864	-0.990

FIGURE 5 | Genetic analysis of the effects of *BnFLC.A2*, *BnFLC.C2*, and *qFT.A3* (*BnFLC.A3b*). m, mean effect; a, additive effect; d, dominant effect; others, interaction effects; the footnotes *A2*, *BnFLC.A2*; *C2*, *BnFLC.C2*; A3, *qFT.A3* (*BnFLC.A3b*); FT, flowering time (days); BN, branch number; BH, branch height (cm); PH, plant height (cm); SH, stem height (cm); ML, main raceme length (cm); MSN, silique number on main raceme; PSN, silique number per plant; PY, plant yield (g); SL, silique length (mm); SSN, seed number per silique; TSW, thousand seed weight (g); OC, oil content (%); PC, protein content (%); GC, glucosinolate content (μmol/g). The red color indicates positive effects, while the blue color indicates negative effects. Best linear unbiased estimation values of the phenotypes for the 15 traits of the 27 NIL genotypes in the Jiangling, Wuhan, and Minle environments were used in the calculation.

seed PC was raised with the postponed FT in the eight genotypes (**Supplementary Figure 3**). In addition, other traits, including ML, SL, and GC, also displayed a significant difference among the eight genotypes, though not exhibiting identical rules (**Supplementary Figure 3**).

Since the above observations clearly demonstrated that the three *BnFLC* homologs greatly contributed to FT variations in NIL backgrounds, the individual genetic effect of these homologs on FT and other traits were dissected by the additive-epistasis model *via* the BLUE values of the eight genotypes (**Supplementary Table 3**). The results showed that *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* exerted the positive additive effects on FT, among which the additive effect of *BnFLC.A2* is greater than that of *BnFLC.C2* and *BnFLC.A3b*

TABLE 2 | The Pearson correlation coefficients between flowering time and other traits.

FT flowering time	BN branch number	BH branch height	PH plant height	SH stem height	ML main raceme length	MSN silique number on main raceme	PSN silique number per plant	PY plant yield	SL silique length	SSN seed number per silique	TSW thousand seed weight	OC oil content	PC protein content	GC glucosinolate content
Pearson correlation coefficients (the eight genotypes)	0.89**	0.95**	0.99**	0.97**	0.055	0.95**	0.84**	0.98**	0.63	0.82*	−0.42	0.74*	−0.85**	−0.13
Pearson correlation coefficients (the 19 NIL hybrids)	0.93**	0.98**	0.99**	0.98**	−0.32	0.96**	0.78**	0.89**	0.86**	0.86**	−0.70**	0.89**	−0.91**	−0.017
Pearson correlation coefficients (the 27 NIL genotypes)	0.9**	0.97**	0.97**	0.97**	−0.16	0.93**	0.58**	0.76**	0.74**	0.84**	−0.59**	0.82**	−0.88**	−0.062
Pearson correlation coefficients (the 16 non-NIL hybrids)	0.55*	0.96**	0.87**	0.93**	−032	0.51*	0.42	0.72**	0.25	0.64**	0.12	0.20	−0.46	0.082

**significant at $p < 0.05$; **significant at $p < 0.01$.*

(**Figure 5**), suggesting that *BnFLC.A2* may be a functional *FLC* homolog with a stronger influence on FT than the other two in rapeseed. However, the additive × additive and additive × additive × additive effects between these three homologs were negative, indicating dose effects among the three genes (**Figure 5**).

Similar to FT, positive additive effects that were also detected on plant architecture traits BN, BH, PH, and SH, and *BnFLC.A2* showed higher additive effects than *BnFLC.C2* and *BnFLC.A3b* in these traits (**Figure 5**). For seed quality traits, positive additive effects were detected on OC and negative additive effects were detected on PC, and *BnFLC.C2* showed stronger effects on OC and PC than *BnFLC.A2* and *BnFLC.A3b* (**Figure 5**). As for GC, it was not affected by the additive effect of any of the three genes, which coincided with its null correlation with FT (**Figure 5** and **Table 2**). For yield component traits, *BnFLC.A2* and *BnFLC.A3b* exerted higher positive additive effects on MSN and PSN than *BnFLC.C2*, coincided with the less silique number phenotype of ID02-010 (**Figure 5** and **Supplementary Table 3**), while *BnFLC.C2* and *BnFLC.A3b* exerted greater positive additive effects on SL and SSN than *BnFLC.A2*, which corresponded with the longer SL and higher SSN of genotype ID05-011 (**Figure 5** and **Supplementary Table 3**). In addition, *BnFLC.A2* and *BnFLC.A3b* exerted negative additive effects on TSW (**Figure 5**). At last, the additive effects of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* on PY were 1.446, 1.365, and 1.361 g, respectively, coincided with the significant Pearson correlation coefficients of FT and PY (**Table 2**), which indicated the acceleration of FT was associated with a severe yield loss, while the delay of FT was associated with yield gain (**Figure 5**).

Heterozygous Status in *BnFLC.A2, BnFLC.C2, and BnFLC.A3b* Generated Accumulative Heterosis in Plant Yield *via* Nonadditive Effects

As described above, there is an obvious conflict between FT and PY in homozygous NIL backgrounds. To test whether this conflict remained in their hybrids, we mutually crossed the eight genotypes and constructed a total of 19 NIL hybrids with heterozygous genotypes at least at one locus of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*, corresponding to genotypes from ID09-020 to ID027-121 (**Supplementary Table 4**). We investigated the FT and PY of these hybrids and performed multiple comparisons of both traits. Abundant variations in FT and PY were retained in different hybrids, ranging from 70.40 to 115.1 days, and from 18.83 to 26.78 g, respectively (**Supplementary Figure 4**). Compared to the original eight homozygous genotypes, the difference of FT in hybrid NILs was narrowed down, whereas that of PY was amplified and predisposed to the high-value end (**Supplementary Figure 4**). Consistently, the Pearson correlation coefficient between the PY and FT reduced to 0.76 when the 19 NIL hybrids joined into the analysis, rather smaller than that 0.98 in the eight homozygous genotypes (**Table 2**).

According to the FT phenotypes, the hybrids can be classified into two groups. In Group I, the FT of hybrids was lower or had

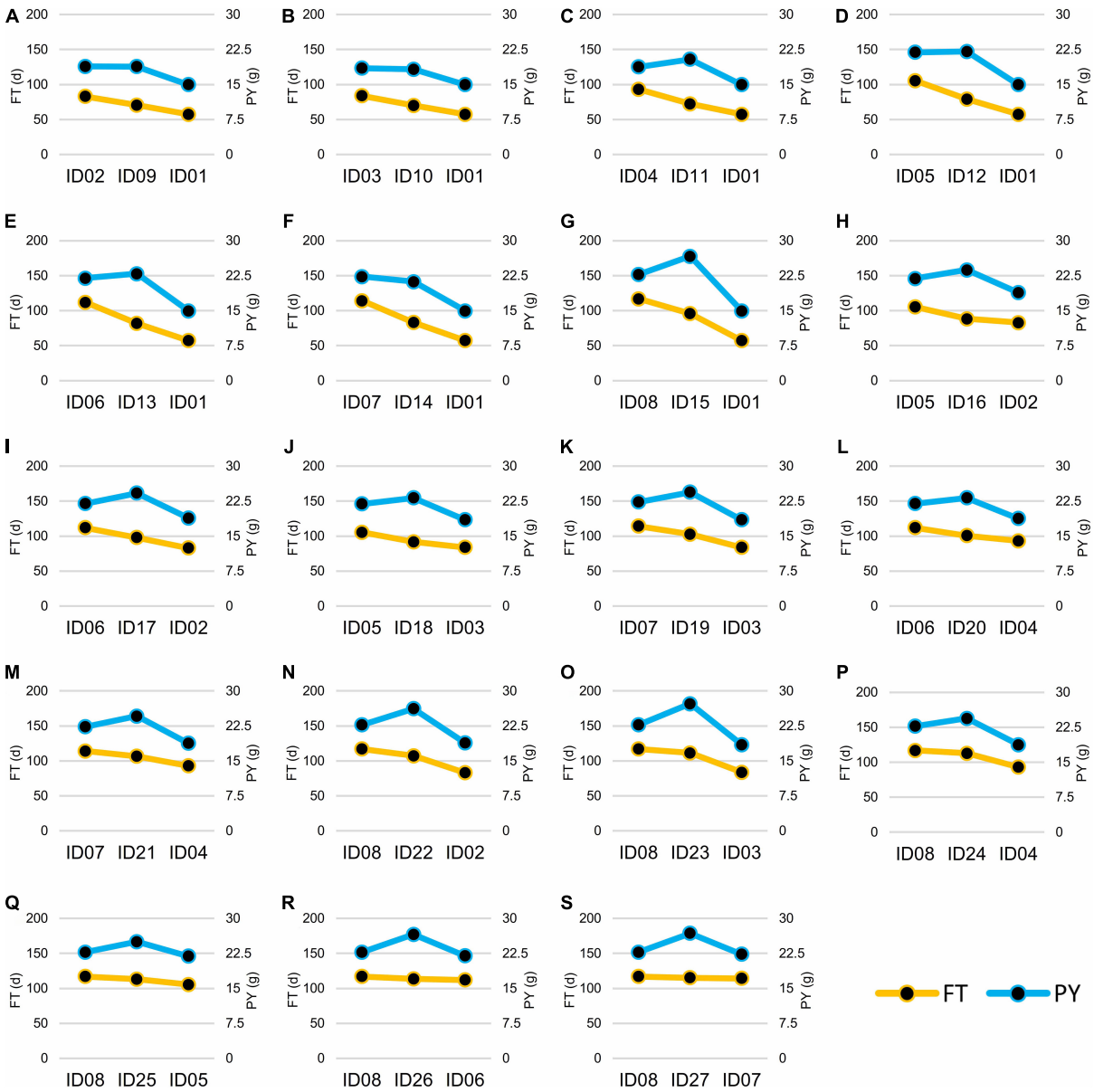


FIGURE 6 | Flowering time (FT) and plant yield (PY) of the 19 NIL hybrids and their crossing parents. For each of the box from **(A–S)**, the left side was the female parent, the right side was the male parent, and the middle was the hybrid. The left vertical axis and the yellow line showed FT (d), while the right vertical axis and the blue line showed PY (g).

no significant difference than the mid-parent value. In Group II, the FT of hybrids was significantly higher than the mid-parent value (**Figure 6**). In Group I, 12 of the 16 hybrids showed no significance of PY to the high-value parents of PY, except for ID16-012, IDh18-021, IDh26-112, and IDh-121, whose PY was significantly higher than their high-value parents (**Figure 6**). In Group II, 6 of the 7 hybrids showed significantly higher PY than the high-value parents, except for ID24-122, whose PY showed no significance to its high-value parent (**Figure 6**). In addition, we noticed that the mid-parent heterosis for plant architecture traits and yield component traits was also observed in these hybrids (**Table 3**). Among them, PSN and PY showed stronger heterosis performances than the other traits (**Table 3**), suggesting the heterosis of PSN was the most important component of the heterosis of PY. Finally, the highest heterosis was found in the triple heterozygous genotype ID15-222 from group II, reaching 25.05 and 41.49% for PSN and PY, respectively (**Table 3**), indicating the importance of the increasing of the heterozygosity for the heterosis of PSN and PY. These findings suggested that in the NIL background, FT was largely inherited as an additive manner, while the PY was consistently inherited as a dominant manner. As a result, the conflict between early FT and high PY

TABLE 3 | Analysis of mid-parent heterosis of the 19 NIL hybrids.

Mid-parent heterosis of the hybrids	FT Flowering time	BN Branch number	BH Branch height	PH Plant height	SH Stem height	ML Main raceme length	MSN Silique number on main raceme	PSN Silique number per plant	PY Plant yield	SL Silique length	SSN Seed number per silique	TSW Thousand seed weight	OC Oil content	PC Protein content	GC Glycosylate content
ID09-020	0.33%	5.61%	7.76%	3.52%	6.95%	−0.12%	8.79%	7.47%	11.35%	0.66%	−5.89%	0.48%	−1.05%	2.53%	1.35%
ID10-002	−1.02%	−1.45%	−5.17%	1.25%	0.86%	1.59%	0.15%	6.95%	9.19%	0.37%	−2.26%	1.45%	−1.13%	2.43%	−3.25%
ID11-200	−3.97%	−4.83%	−8.69%	1.95%	−3.47%	7.56%	3.64%	14.27%	21.03%	−0.04%	0.70%	1.02%	1.28%	−1.47%	2.84%
ID12-022	−2.92%	−1.80%	1.86%	2.21%	−3.16%	9.07%	1.47%	18.44%	19.76%	−5.20%	−9.78%	2.07%	−0.83%	0.89%	−1.84%
ID13-220	−3.53%	3.92%	−3.98%	4.87%	−0.71%	12.23%	2.50%	17.84%	24.19%	−1.44%	−3.06%	-0.53%	2.61%	−5.89%	−2.32%
ID14-202	−3.06%	−6.52%	−12.02%	1.29%	−10.34%	18.08%	6.77%	12.32%	13.85%	−1.27%	−8.62%	2.32%	0.64%	−1.42%	1.31%
ID15-222	9.89%	2.27%	16.45%	11.31%	7.09%	17.56%	19.40%	25.05%	41.49%	3.38%	0.00%	0.48%	4.44%	−5.64%	−1.82%
ID16-012	−6.15%	10.22%	3.93%	2.42%	5.89%	−2.32%	2.14%	18.51%	16.41%	−3.11%	−2.03%	−2.85%	−2.22%	4.02%	−2.32%
ID17-210	0.28%	12.80%	0.30%	3.48%	8.09%	−2.98%	8.70%	18.13%	18.61%	1.06%	0.03%	−3.39%	−1.30%	3.31%	−0.34%
ID18-021	−2.97%	−0.11%	−4.09%	−1.01%	−3.27%	2.04%	−1.82%	17.69%	14.73%	−3.14%	−7.42%	−1.10%	−3.04%	6.35%	−8.23%
ID19-201	3.46%	10.39%	2.40%	4.78%	6.01%	2.84%	10.86%	16.02%	19.55%	2.07%	−2.18%	3.62%	1.85%	−0.58%	−4.27%
ID20-120	−1.88%	5.66%	2.15%	3.50%	5.38%	0.82%	5.26%	12.35%	13.80%	2.37%	−5.08%	1.51%	−2.29%	5.45%	0.98%
ID21-102	2.97%	8.55%	16.64%	7.71%	12.90%	−0.08%	8.00%	11.01%	19.73%	3.28%	−2.71%	3.11%	-0.29%	1.46%	0.12%
ID22-212	7.30%	8.93%	16.54%	10.62%	13.10%	6.71%	16.23%	14.68%	26.03%	7.13%	3.87%	4.45%	2.06%	−1.04%	5.91%
ID23-221	10.84%	1.51%	26.23%	13.02%	19.79%	2.30%	13.63%	16.56%	32.24%	5.88%	13.05%	0.35%	2.97%	−1.25%	−3.89%
ID24-122	7.53%	5.99%	30.29%	11.21%	19.21%	−1.15%	12.55%	10.45%	17.66%	3.57%	6.35%	0.92%	3.62%	−2.64%	−0.37%
ID25-211	1.68%	3.47%	18.36%	6.95%	10.72%	0.12%	3.46%	9.82%	12.02%	−1.85%	1.80%	−4.97%	0.62%	−2.40%	4.33%
ID26-112	−0.76%	11.11%	2.90%	6.18%	8.51%	1.84%	8.20%	16.52%	18.81%	0.48%	2.03%	−4.77%	0.91%	−1.89%	0.15%
ID27-121	−0.47%	8.58%	5.57%	5.13%	8.27%	−1.12%	2.92%	15.03%	18.91%	−0.67%	1.20%	−0.97%	2.21%	−2.65%	2.02%

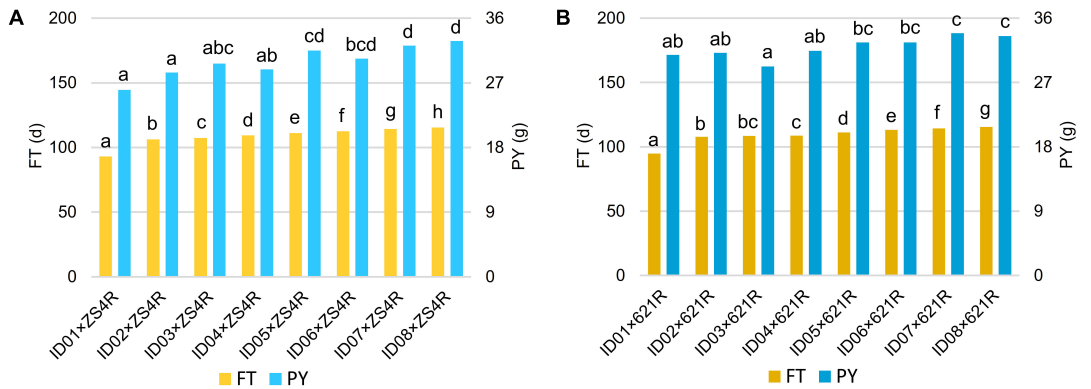


FIGURE 7 | Multiple comparisons of FT and PY of the non-NIL hybrids. **(A)** Multiple comparisons of the hybrids produced by crossing with ZS4R. **(B)** Multiple comparisons of the hybrids produced by crossing with 621R. The left bars indicate FT (d), and the right bars indicate PY (g). The left legend stands for FT (d), and the right legend stands for PY (g), and the lower-case letters next to the bars showed the significance of multiple comparisons.

can be greatly overcome by the heterosis derived from *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* in the hybrid-NIL backgrounds.

To support the above conclusion, we further dissected the genetic effects of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* on FT, PY, and PY component traits by the BLUE values of the 27 NIL genotypes using the additive–dominant–epistasis model (**Supplementary Table 3**). The results showed that there are very weak nonadditive effects ranging from –2.254 to 5.297 days on FT existed for the three *FLC* homologs (**Figure 5**), indicating less mid-parent heterosis for FT. In contrast, *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* had much stronger dominant effects on PY (3.504, 2.991, and 3.284 g, respectively) than their additive effects (1.446, 1.365, and 1.361 g, respectively; **Figure 5**), indicating the dominant effects were the important reason for PY heterosis (**Figure 5** and **Table 3**). We also noticed that the dominant × dominant × dominant effect of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* on PY was 4.393 g, a huge interaction for the gain of PY, which was coincided with the highest mid-parent heterosis and the highest PY of ID15-222. Therefore, we confirmed that the genotype which was heterozygous on *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* could simultaneously be the most preponderant genotype in the 27 NIL genotypes for PY heterosis.

Heterosis in Plant Yield Is Also Associated With Heterozygous *BnFLC.A2, BnFLC.C2,* and/or *BnFLC.A3b* Genotypes in Non-NIL Hybrids

To test whether the heterosis of PY mediated by *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* can be explored in early-flowering hybrid breeding, we investigated the FT and PY of 16 non-NIL hybrids which were constructed by crossing two commercial pol CMS restorers (i.e., ZS4R and 621R) to the eight genotypes (i.e., ID01–ID08). Genotyping by PCR analysis showed that ZS4R carries the same alleles in the *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* loci as 616A (*Bnflc.a2/Bnflc.c2/BnFLC.A3b*), thus designated as 110, while 621R has the late-flowering alleles in all the three genes (*BnFLC.A2/BnFLC.C2/BnFLC.A3b*), thus designated as 111. As shown in **Figure 7**, we observed that in both hybrid groups, the resulting hybrids exhibited significant differences in FT and displayed a similar tendency as the eight genotypes (**Figure 4** and **Supplementary Figure 3**), varying from 93.05 to 115.5 days for the ZS4R hybrid group and from 9.65 to 115.6 days for the 621R hybrid group, respectively. Accordingly, these results indicate that the different alleles and their permutations also have a significant and identical effect on FT in non-NIL hybrids. Consistently, it can easily be noticed that with the extension of FT, the PY performance was also gradually increased in the non-NIL hybrids (**Figures 7A,B**), with a significant Pearson correlation coefficient of 0.72 (**Table 2**), showing that early flowering caused yield loss in non-NIL hybrids as well.

However, we also observed that the heterozygous status of the *BnFLC* homologs successfully made up the yield reduction in some hybrids. For example, ID06 × ZS4R-110 from the ZS4R hybrid group, homozygous in all the *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* genes, had an FT of 112.5 days, which was significantly later than ID02 × ZS4R-210, ID03 × ZS4R-222, ID04 × ZS4R-120, and ID05 × ZS4R-212, while there was no significant difference in PY between ID06 × ZS4R-110 and the other four hybrids (**Figure 7A**). Consistently, it was also clearly revealed that ID01 × 621R-222 from the 621R hybrid group, which is heterozygous in all the *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* genes, showed the earliest FT phenotype among all the hybrids in the same group, while ID01 × 621R-222 displayed the same PY level as ID02 × 621R-212, ID03 × 621R-221, ID04 × 621R-122, ID05 × 621R-211, and ID06 × 621R-112, suggesting that a 12–17-day FT difference did not cause obvious yield loss in these hybrids (**Figure 7B**). Similarly, ID05 × 621R-211, ID06 × 621R-112, and ID07 × 621R-121 had no significant difference in PY with ID08 × 621R-111, which is homozygous in all the *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* genes, but their FT was 1.2–4.4 days earlier than that of ID08 × 621R-111, respectively (**Figure 7B**). In summary, all these results showed that the exploration of the heterosis of *BnFLC* homologs

in complex genetic background can effectively rescue the PY damage during early-flowering breeding as well.

DISCUSSION

In this study, we discovered *qFT.A3* as an early flowering factor of pol CMS line 616A by using a QTL mapping strategy. According to the latest genome information of ZS11 (Song et al., 2020a), *BnFLC.A3b* is the only annotated FT gene in the 727 kb candidate region of *qFT.A3*. Similarly, *BnFLC.A3b* was previously colocalized with an FT QTL located on chromosome A3 from a DH population that derived the F_1 cross of Ningyou7 and Tapidor (Wang et al., 2011; Zou et al., 2012). *BnFLC.A3b* is expressed normally in the winter-type varieties, such as Darmor, Tapidor, and Quinta, but is considerably less expressed in the spring- or semi-winter-type varieties, such as Westar, ZS11, and Ningyou7 (Zou et al., 2012; Song et al., 2020b; Yin et al., 2020). Alternative splicing induced by intron variation was probably responsible for the differential expression of *BnFLC.A3b* between Ningyou7 and Tapidor, hinting that *BnFLC.A3b* was thus presumed to be a functional FT gene (Zou et al., 2012). Here, we also identified rich InDel and SNP variations in *BnFLC.A3b* between R11 and 616A, and observed a 13-fold higher expression level of *BnFLC.A3b* in the former parents than the latter. Therefore, even with the mutant alleles *Bnflc.a2* and *Bnflc.c2*, which greatly promote early flowering, R11 displayed a significant late-flowering phenotype than 616A in the winter environment. All these suggest that *BnFLC.A3b* is the most likely target gene of *qFT.A3*, which acts as a key FT promoting factor even under the existence of functional *BnFLC.A2* and *BnFLC.C2*.

The rational design of FT enables breeders to adapt varieties to various environments. As a subtropical region, central China is suitable for semi-winter rapeseed, indica rice, and maize cultivation. "Rice—ratoon rice—rapeseed," "double season rice—rapeseed," and "maize—rapeseed" farming systems are being promoted in this region to simultaneously produce rice, corn, and rapeseed. However, the limitations of growth season require rapeseed to be sowed later and harvested earlier for saving time for rice or corn, because rice and corn are more important in China (Xu et al., 2018; Zhao and Yang, 2018; Yuan et al., 2019). As a warm temperate region, northwestern China is suitable for one season spring rapeseed, in which rapeseed needs to be harvested earlier for escaping the early frost. In this study, we employed natural variants of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* to create sterile lines with the 616A background in an effort to adapt rapeseed hybrids to different farming systems and environments.

As flowering is the critical transition from vegetative to reproductive growth, besides FT, FT genes can affect many other agronomic traits. Among the traits affected in this study, the FT genes did not only have additive effects on PY and plant architecture component traits but also had dominant effects on these traits. Smaller plant architecture corresponds to lower dry matter production, indicating that a correlation between early FT and yield loss may be an indirect reflection of plant architecture and dry matter production (Miedaner et al., 2018). Suggesting the yield loss induced by early flowering was a comprehensive reflection in the improvement of FT.

The heterosis or nonadditive effects of FT genes afforded us an opportunity to balance yield when FT was advanced, but this also complicated the influence of FT on other agronomic traits. In this study, the nonadditive effects of *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b* successfully increased heterosis, providing an additional example of FT gene heterosis as well as an innovative germplasm for overcoming the conflict between early flowering and yield loss in rapeseed breeding. In previous studies, heterotic QTLs *Ghd7* and *Ehd3* resulted from the introgression of subpopulation alleles and diversity selection in rice (Lin et al., 2020). Yield heterosis in tomatoes resulting from heterozygous mutations of *SFT* (*sft*/+) was hypothesized to be based on altered amounts of functional SFT proteins within each modular sympodial unit (Krieger et al., 2010; Park et al., 2014). Response of genes to carbohydrate stimulus and nutrient levels might be involved in the generation of the large plant phenotype of F_1 mimics in *Arabidopsis*. In the current study, the heterosis of the *BnFLC* homologs might be embodied through downstream *FT* paralogs, similar as in tomatoes. We will pursue research investigating the mechanism of FT gene heterosis in rapeseed in the future.

In this study, the elite line 616A is the female parent of several registered cultivars, including Shengguang136 (616A crossed with ZS4R) and Shengguang 128 (616A crossed with 621R). Shengguang 136 is suitable for the climate of central China, and Shengguang 128 is adaptable to both environments of central China and northwestern China, but more maturity types are still required. In the Jiangling and Wuhan environments representing the climate in central China, IDh61-222 (ID01-000 × 621R) was shown to be appropriate for commercial use because of its suitable early flowering and harvest times. In the Minle environment, which represents the climate in northwestern China, IDh67-121 (ID07-101 × 621R) and IDh68-111 (ID06-110 × 621R) were suggested to gain a higher yield without wasting light and temperature conditions. In the future, the eight genotypes of male sterile lines of 616A will be crossed with additional restore lines and tested in more environments to breed hybrids adaptable to different farming systems in different sub-environments (**Supplementary Figure 5**).

CONCLUSION

In this study, we dissected the genetics of the early FT of the elite lines 616A and R11, and classified that yield loss induced by early flowering could be compensated by the heterosis of FT genes *BnFLC.A2*, *BnFLC.C2*, and *BnFLC.A3b*. These results not only indicated that the dominant effect is an important contributor to heterosis but also supported a feasible method to facilitate yield *via* heterosis of FT genes during the rational design of elite cultivars suitable for the current farming systems and different sub-environments. Moreover, the results suggested that the sterile lines of the eight genotypes developed in the current study could be potential for hybrid breeding in the near future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The sequences of BnFLC.A3b-616A and BnFLC.A3b-R11 were deposited at GenBank (https://www.ncbi.nlm.nih.gov/) with the accession numbers of OL674083 and OL674084, respectively.

AUTHOR CONTRIBUTIONS

CF, GY, DH, and FD conceived and designed the experiments. CF, ZW, YS, and AA performed the experiments. CF and PW analyzed the data. CF, GY, and DH wrote this manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the China Agriculture Research System of MOF and MARA (CRS12), the Open Fund of the National Key Laboratory of Crop Genetic Improvement (ZK201909), the National Natural Science Foundation of China (31401417), and the Wuhan Applied Foundational Frontier Project (2019020701011446).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.798371/full#supplementary-material

REFERENCES

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384

Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* 44, D1167–D1171. doi: 10.1093/nar/gkv1054

Chen, L., Dong, F., Cai, J., Xin, Q., Fang, C., Liu, L., et al. (2018). A 2.833-kb insertion in BnFLC.A2 and its homeologous exchange with BnFLC.C2 during breeding selection generated early-flowering rapeseed. *Mol. Plant* 11, 222–225. doi: 10.1016/j.molp.2017.09.020

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Dunne, F. L., Kelleher, M., Walsh, S. W., and Berry, D. P. (2018). Characterization of best linear unbiased estimates generated from national genetic evaluations of reproductive performance, survival, and milk yield in dairy cows. *J. Dairy Sci.* 101, 7625–7637. doi: 10.3168/jds.2018-14529

Guo, Y., Hans, H., Christian, J., and Molina, C. (2014). Mutations in single FT- and TFL1-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Front. Plant Sci.* 5:282. doi: 10.3389/fpls.2014.00282

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

Jiang, P., Wang, S., Jiang, H., Cheng, B., Wu, K., and Ding, Y. (2018). The COMPASS-Like complex promotes flowering and panicle branching in rice. *Plant Physiol.* 176, 2761–2771. doi: 10.1104/pp.17.01749

Kaur, S., Atri, C., Akhatar, J., Mittal, M., Kaur, R., and Banga, S. S. (2021). Genetics of days to flowering, maturity and plant height in natural and derived forms of *Brassica rapa* L. *Theor. Appl. Genet.* 134, 473–487. doi: 10.1007/s00122-020-03707-9

Khan, G. H., Shikari, A. B., Vaishnavi, R., Najeeb, S., Padder, B. A., Bhat, Z. A., et al. (2018). Marker-assisted introgression of three dominant blast resistance genes into an aromatic rice cultivar Mushk Budji. *Sci. Rep.* 8:4091. doi: 10.1038/s41598-018-22246-4

Korostin, D., Kulemin, N. A., Naumov, V. A., Belova, V. À, Kwon, D., and Gorbachev, A. (2020). Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. *PLoS One* 15:e0230301. doi: 10.1371/journal.pone.0230301

Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463. doi: 10.1038/ng.550

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H., Handsaker, R. E., 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., Younas, M., Wang, X., Li, X., Chen, L., Zhao, B., et al. (2013). Development of a core set of single-locus SSR markers for allotetraploid rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 126, 937–947. doi: 10.1007/s00122-012-2027-z

Li, S., Ying, Y., Secco, D., Wang, C., Narsai, R., Whelan, J., et al. (2017). Molecular interaction between PHO2 and GIGANTEA reveals a new crosstalk between flowering time and phosphate homeostasis in *Oryza sativa*. *Plant Cell Environ.* 40, 1487–1499. doi: 10.1111/pce.12945

Lin, Z., Qin, P., Zhang, X., Fu, C., Deng, H., Fu, X., et al. (2020). Divergent selection and genetic introgression shape the genome landscape of heterosis in hybrid rice. *Proc. Natl. Acad. Sci. U.S.A.* 117, 4623–4631. doi: 10.1073/pnas.1919086117

Liu, K., Yu, Y., Dong, A., and Shen, W. H. (2017). SET DOMAIN GROUP701 encodes a H3K4-methytransferase and regulates multiple key processes of rice plant development. *New Phytol.* 215, 609–623. doi: 10.1111/nph.14596

Liu, Z., Liu, P., Long, F., Hong, D., He, Q., and Yang, G. (2012). Fine mapping and candidate gene analysis of the nuclear restorer gene Rfp for pol CMS in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 125, 773–779. doi: 10.1007/s00122-012-1870-2

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

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., et al. (2017). Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* 49, 773–779. doi: 10.1038/ng.3819

Mannai, Y., Akabane, K., Hiratsu, K., Satoh-Nagasawa, N., and Wabiko, H. (2017). The NAC transcription factor gene OsY37 (ONAC011) promotes leaf senescence and accelerates heading time in rice. *Int. J. Mol. Sci.* 18:2165. doi: 10.3390/ijms18102165

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

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Miedaner, T., Haffke, S., Siekmann, D., Fromme, F., Roux, S., and Hackauf, B. (2018). Dynamic quantitative trait loci (QTL) for plant height predict biomass yield in hybrid rye (*Secale cereale* L.). *Biomass Bioenergy* 115, 10–18. doi: 10.1016/j.biombioe.2018.04.001

Miedaner, T., Schulthess, A. W., Gowda, M., Reif, J. C., and Longin, C. F. (2016). High accuracy of predicting hybrid performance of Fusarium head blight resistance by mid-parent values in wheat. *Theor. Appl. Genet.* 130, 461–470. doi: 10.1007/s00122-016-2826-8

Minh-Thu, P. T., Kim, J. S., Chae, S., Jun, K. M., Lee, G. S., Kim, D. E., et al. (2018). A WUSCHEL homeobox transcription factor, OsWOX13, enhances drought tolerance and triggers early flowering in rice. *Mol. Cells* 41, 781–798. doi: 10.14348/molcells.2018.0203

Park, S. J., Jiang, K., Tal, L., Yichie, Y., Gar, O., Zamir, D., et al. (2014). Optimization of crop productivity in tomato using induced mutations in the Florigen pathway. *Nat. Genet.* 46, 1337–1342. doi: 10.1038/ng.3131

Peterson, B. G., and Carl, P. (2020). *Econometric Tools for Performance and Risk Analysis* [R package PerformanceAnalytics version 2.0.4].

Raman, H., Raman, R., Qiu, Y., Yadav, A. S., Sureshkumar, S., Borg, L., et al. (2019). GWAS hints at pleiotropic roles for FLOWERING LOCUS T in flowering time and yield-related traits in canola. *BMC Genomics* 20:636. doi: 10.1186/s12864-019-5964-y

Schiessl, S. (2020). Regulation and subfunctionalization of flowering time genes in the allotetraploid oil crop *Brassica napus*. *Front. Plant Sci.* 11:605155. doi: 10.3389/fpls.2020.605155

Schiessl, S., Samans, B., Hüttel, B., Reinhardt, R., and Snowdon, R. J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 5:404. doi: 10.3389/fpls.2014.00404

Song, J., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020a). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Song, J. M., Liu, D., Xie, W., Yang, Z., Guo, L., Liu, K., et al. (2020b). BnPIR: *Brassica napus* pan-genome information resource for 1689 accessions. *Plant Biotechnol. J.* 19, 412–414. doi: 10.1111/pbi.13491

Song, S., Wang, G., Hu, Y., Liu, H., Bai, X., Qin, R., et al. (2018). OsMFT1 increases spikelets per panicle and delays heading date in rice by suppressing Ehd1, FZP and SEPALLATA-like genes. *J. Exp. Bot.* 69, 4283–4293. doi: 10.1093/jxb/ery232

Wang, N., Qian, W., Suppanz, I., Wei, L., Mao, B., Long, Y., et al. (2011). Flowering time variation in oilseed rape (*Brassica napus* L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. *J. Exp. Bot.* 62, 5641–5658. doi: 10.1093/jxb/err249

Wang, Y., Xu, C., Sun, J., Dong, L., Li, M., Liu, Y., et al. (2021). GmRAV confers ecological adaptation through photoperiod control of flowering time and maturity in soybean. *Plant Physiol.* 187, 361–377. doi: 10.1093/plphys/kiab255

Wang, Z., Wan, L., Zhang, X., Xin, Q., Song, Y., Hong, D., et al. (2021). Interaction between Brassica napus polygalacturonase inhibition proteins and Sclerotinia sclerotiorum polygalacturonase: implications for rapeseed resistance to fungal infection. *Planta* 253:34. doi: 10.1007/s00425-020-03556-2

Xi, X., Wei, K., Gao, B., Liu, J., Liang, J., Cheng, F., et al. (2018). BrFLC5: a weak regulator of flowering time in *Brassica rapa*. *Theor. Appl. Genet.* 131, 2107–2116. doi: 10.1007/s00122-018-3139-x

Xu, L., Zhan, X., Yu, T., Nie, L., Huang, J., Cui, K., et al. (2018). Yield performance of direct-seeded, double-season rice using varieties with short growth durations in central China. *Field Crops Res.* 227, 49–55. doi: 10.1016/j.fcr.2018.08.002

Yin, S., Wan, M., Guo, C., Wang, B., Li, H., Li, G., et al. (2020). Transposon insertions within alleles of BnaFLC.A10 and BnaFLC.A2 are associated with seasonal crop type in rapeseed. *J. Exp. Bot.* 71, 4729–4741. doi: 10.1093/jxb/eraa237

Yuan, S., Cassman, K. G., Huang, J., Peng, S., and Grassini, P. (2019). Can ratoon cropping improve resource use efficiencies and profitability of rice in central China? *Field Crops Res.* 234, 66–72. doi: 10.1016/j.fcr.2019.02.004

Yuan, Y. X., Jian, W., Sun, R. F., Zhang, X. W., Xu, D. H., Bonnema, G., et al. (2009). A naturally occurring splicing site mutation in the *Brassica rapa* FLC1 gene is associated with variation in flowering time. *J. Exp. Bot.* 60, 1299–1308. doi: 10.1093/jxb/erp010

Zhang, H., Zhu, S., Liu, T., Wang, C., Cheng, Z., Zhang, X., et al. (2019). DELAYED HEADING DATE1 interacts with OsHAP5C/D, delays flowering time and enhances yield in rice. *Plant Biotechnol. J.* 17, 531–539. doi: 10.1111/pbi.12996

Zhao, J., and Yang, X. (2018). Distribution of high-yield and high-yield-stability zones for maize yield potential in the main growing regions in China. *Agric. For. Meteorol.* 248, 511–517. doi: 10.1016/j.agrformet.2017.10.016

Zhu, Y. J., Fan, Y. Y., Wang, K., Huang, D. R., Liu, W. Z., Ying, J. Z., et al. (2017). Rice Flowering Locus T 1 plays an important role in heading date influencing yield traits in rice. *Sci. Rep.* 7:4918. doi: 10.1038/s41598-017-05302-3

Zou, X., Suppanz, I., Raman, H., Hou, J., Wang, J., Long, Y., et al. (2012). Comparative analysis of FLC homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. *PLoS One* 7:e45751. doi: 10.1371/journal.pone.0045751

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 in Plant Science

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fpls.2022.801456

Site-Directed Mutagenesis of the Carotenoid Isomerase Gene *BnaCRTISO* Alters the Color of Petals and Leaves in *Brassica napus* L.

Huailin Li[1,2†], Kaidi Yu[1,2†], Olalekan Amoo[1,2], Yalun Yu[1,2], Mixia Guo[1,2], Songyue Deng[1,2], Mengting Li[1,2], Limin Hu[1,2], Jingzhen Wang[1,2], Chuchuan Fan[1,2] and Yongming Zhou[1,2]*

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

OPEN ACCESS

Edited by:
Kun Lu,
Southwest University, China

Reviewed by:
Fuyou Fu,
Agriculture and Agri-Food Canada (AAFC), Canada
Sun-Hwa Ha,
Kyung Hee University, South Korea

***Correspondence:**
Chuchuan Fan
fanchuchuan@mail.hzau.edu.cn

[†] These authors have contributed equally to this work

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

Received: 25 October 2021
Accepted: 03 January 2022
Published: 10 February 2022

Citation:
Li H, Yu K, Amoo O, Yu Y, Guo M, Deng S, Li M, Hu L, Wang J, Fan C and Zhou Y (2022) Site-Directed Mutagenesis of the Carotenoid Isomerase Gene BnaCRTISO Alters the Color of Petals and Leaves in Brassica napus L. Front. Plant Sci. 13:801456. doi: 10.3389/fpls.2022.801456

The diversity of petal and leaf color can improve the ornamental value of rapeseed and promote the development of agriculture and tourism. The two copies of carotenoid isomerase gene (*BnaCRTISO*) in *Brassica napus* (*BnaA09.CRTISO* and *BnaC08.CRTISO*) was edited using the CRISPR/Cas9 system in the present study. The mutation phenotype of creamy white petals and yellowish leaves could be recovered only in targeted mutants of both *BnaCRTISO* functional copies, indicating that the redundant roles of *BnaA09.CRTISO* and *BnaC08.CRTISO* are vital for the regulation of petal and leaf color. The carotenoid content in the petals and leaves of the *BnaCRTISO* double mutant was significantly reduced. The chalcone content, a vital substance that makes up the yellow color, also decreased significantly in petals. Whereas, the contents of some carotenes (lycopene, α-carotene, γ-carotene) were increased significantly in petals. Further, transcriptome analysis showed that the targeted mutation of *BnaCRTISO* resulted in the significant down-regulation of important genes *BnaPSY* and *BnaC4H* in the carotenoid and flavonoid synthesis pathways, respectively; however, the expression of other genes related to carotenes and xanthophylls synthesis, such as *BnaPDS3*, *BnaZEP*, *BnaBCH1* and *BCH2*, was up-regulated. This indicates that the molecular mechanism regulating petal color variation in *B. napus* is more complicated than those reported in *Arabidopsis* and other *Brassica* species. These results provide insight into the molecular mechanisms underlying flower color variation in rapeseed and provides valuable resources for rapeseed breeding.

Keywords: ***Brassica napus*, flower color, *BnaCRTISO*, gene editing, carotenoid**

INTRODUCTION

Flower color is considered one of the major attractants for pollen transmission in nature. Insects can recognize different flower colors through visual signals and transmit targeted pollen between flowers (Cazzonelli, 2011; Nisar et al., 2015). It is generally recognized that flower color is determined by three main pigments: carotenoids, flavonoids, and betalains (Grotewold, 2006).

Carotenoids mainly provide orange and red color, flavonoids provide yellow, and betalains provide a few other colors (Cazzonelli, 2011). They are fat-soluble terpenoids synthesized *via* the isoprene pathway and have been found in various plants and animals (Cunningham and Gantt, 1998). Carotenoids are important pigments determining the color of fruits, vegetables, and flowers (Lv et al., 2015; Zhang et al., 2015; Li X. et al., 2018). They are found mainly in leaf, flower, fruit, and root tissues and play a vital role during plant development, such as protecting the plants against photo-oxidative damage (Holt et al., 2005; Sun et al., 2018). Carotenoids are also the precursors for the biosynthesis of vitamin A, the plant hormones abscisic acid (ABA) and strigolactone (Walter and Strack, 2011).

The biosynthetic pathway of carotenoid in higher plants has been elucidated, and many genes encoding key enzymes involved in this pathway have been successfully cloned. Based on the presence or absence of oxygen in their molecular structure, carotenoids can be divided into two categories, i.e., carotenes and xanthophylls. Carotenes contain carbon and hydrogen in their molecular structure, such as phytoene, lycopene, and α-carotene; xanthophylls contain oxygen, such as lutein, zeaxanthin, neoxanthin (Montserrat et al., 2015). In general, carotenoids in plants use GGPP (Geranylgeranyl pyrophosphate) as the synthetic substrate except for a few plants (Fraser and Bramley, 2004; Matsufuji et al., 2007). In *Arabidopsis thaliana*, the biosynthesis of carotenoid is regulated by ten enzymes and eleven genes (Patrick et al., 2018).

Carotenoid isomerase (CRTISO) converts the yellow colored prolycopene into the red colored all-*trans* lycopene in the carotenoid synthesis pathway (Breitenbach and Sandmann, 2005). Previous studies have shown that down-regulation of *CRTISO* gene expression in citrus can promote the accumulation of β-carotene (Kato et al., 2007). It was also reported that the expression of *CRTISO* gene could increase carotenoid accumulation in the endosperm of maize seeds (Wurtzel, 2009), implying that the expression of *CRTISO* gene has a certain correlation with the accumulation of carotenoids in plants. At present, the *CRTISO* gene has been cloned in tomato (Isaacson, 2002), maize (Wurtzel, 2009), melon (Galpaz et al., 2013), and *A. thaliana* (Hyoungshin, 2002). But there is no relevant research report on this gene in rapeseed.

Rapeseed is the third-largest oilseed crop worldwide with multi-function application, which are widely used as edible vegetable oil, vegetables, fodder, biofuel and nectar, improvement of saline and alkaline soils (Liu et al., 2020). The distinctive flower color of rapeseed also shows high ornamental value, which has received increased attention in China. The typical flower color of rapeseed is yellow, but there are also reports of some variant flower colors, such as white, yellowish and orange. Numerous genetic analysis of flower color have indicated that this trait is little affected by environmental factors and shows dominant or incompletely dominant inheritance (Quazi, 1988; Zhang et al., 2010; Huang et al., 2014). Up to date, several genes controlling flower color have been reported in *B. napus*. The white flower color is controlled by a single dominant gene, *BnaC3.CCD4*, which encodes a carotenoid cleavage dioxygenase and is involved in carotenoid degradation (Zhang et al., 2015; Han et al., 2019). The insertion of a CACTA-like transposable element in *BnaC3.CCD4* leads to its loss-of-function and a subsequently enhanced accumulation of carotenoids; thus, results in a petal color transition from white to yellow (Zhang et al., 2015; Han et al., 2019). Gene silencing of two genes, *BnaA09.ZEP* and *BnaC09.ZEP*, confers the change in flower color from yellow to orange (Liu et al., 2020). These two genes are homologous to the nuclear-encoded plastid enzyme zeaxanthin epoxidase (AtZEP) and participate in carotenoid biosynthesis. Recently, Zhao et al. (2021) reported that a yellowish-white flower trait is controlled by a single recessive gene, *BnaA08.PDS3*, which encodes a phytoene desaturase 3 (Zhao et al., 2021). A C-to-T substitution in the coding region of *BnaA08.PDS3* results in a premature translation termination and a subsequent decreased carotenoid biosynthesis; thus, changing the flower color from yellow to yellowish-white. Other researchers used cell fusion technology to obtain the fusion plant with white flower trait (47 chromosomes). When the chromosomes were reduced to 38 by backcrossing, the petal color changed from white to yellow (Sakai, 1995). Therefore, until now, the genes and molecular mechanisms regulating flower color in rapeseed have not been fully elucidated.

In recent years, sequence-specific nucleases (SSNs) have been demonstrated to be an amazing tool for improving crops *via* site-specific genome editing, and CRISPR/Cas9 is considered the most simple and efficient SSN. The CRISPR/Cas9 system has been effectively utilized in rapeseed to produce the targeted mutations for the improvement of numerous agronomic traits (Braatz et al., 2017; Yang et al., 2017, 2018; Hu et al., 2018; Li C. et al., 2018; Zhai et al., 2019; Ahmar et al., 2021).

Hence, we utilized the CRISPR/Cas9 system to generate efficient knockouts of *CRTISO* homeologs with stable transformation in rapeseed. In the T_1, T_2, and T_3 generations, mutants containing the desired gene modification were obtained by segregation. The transcriptomic analysis and metabolite profiling of *BnaCRTISO* mutant plants were used in this current study to investigate the molecular mechanisms that regulate the petal color in *B. napus*. This study provided valuable germplasm resources for the innovation of different petal color varieties in rapeseed and offered a new way to improve polypoid crops.

MATERIALS AND METHODS

Plant Materials

In this study, *B. napus* pure line J9707 was used as the donor plants for transformation, and the seeds were obtained from the National Engineering Research Center of Rapeseed, Wuhan, China. The flowers on the primary inflorescence were marked at anthesis, and the petals with different colors were collected for transcriptomic analysis and metabolite profiling.

Construction of the CRISPR/Cas9 Vector and Plant Transformation

The binary pYLCRIPSR/Cas9 multiplex genome targeting vector system was utilized for gene editing in this study (Ma et al., 2015). The selection of sequence-specific sgRNAs in the target gene, CRISPR/Cas9 construct assembly, and *Agrobacterium*

tumefaciens-mediated hypocotyl transformation in *B. napus* were conducted as previously described (Hu et al., 2018). The oligos employed in constructing the sgRNA vectors are listed in **Supplementary Table 1**. The resulting construct is described in detail in **Figure 1B**.

Identification of Transgenic Plants and Potential Off-Targets

The transgenic plants were screened by hygromycin selection (25 mg/L). Then, the presence of the T-DNA in the construct was assessed by PCR using the specific primer pairs PB-L/PB-R (**Supplementary Table 1**).

The targeted mutations were determined in transgenic plants using the high-throughput tracking of mutations (Hi-TOM) platform (Zhai et al., 2019). Target-specific and barcoding PCR (two rounds of PCR) were performed to amplify the genomic region encompassing the specific targets of each independent sample, and the resulting PCR products were mixed in equal amounts and purified for next-generation sequencing (the Illumina HiSeq platform at the Novogene Bioinformatics Institute, Beijing, China). The sequencing data was then decoded using a corresponding online tool to track the mutations of the target sites[1]. The potential off-target sites were identified using CRISPR-P2.0[2]. The primers used to detect targeted and potential off-target mutations are listed in **Supplementary Table 1**.

RNA Extraction and Quantitative Real-Time PCR

Total RNA was prepared using the EasyPure Plant RNA Kit (TransGen Biotech, Beijing, China), and cDNA was synthesized using the Transcript RT Kit (TransGen Biotech). The qRT-PCR was carried out using the TransStart Top Green qPCR SuperMix Kit (TransGen Biotech) on a CFX384 Real-Time System (Bio-Rad). Relative quantification was performed using the comparative cycle threshold method. The relative amount of PCR product that was amplified using the designed primer sets (**Supplementary Table 1**) was normalized to the reference genes, *BnaACT2* and *BnaUBC9*.

RNA-Seq Transcriptomic Analysis

Flower tissues were sampled with three biological replicates. Petals were gently hand dissected from the flower on dry ice, immediately frozen immediately in liquid nitrogen, and stored at −80°C until total RNA extraction.

RNA extraction, cDNA library construction, sequencing, quality control, and read mapping to the reference genome, identification of differentially expressed genes (DEGs), and GO and KEGG pathway enrichment analysis of DEGs were performed using previously established procedure (Shahid et al., 2019; Zhai et al., 2020). Fragments per kilobase of transcript per million mapped reads (FPKM) were calculated as a measure of the level of gene expression. Genes with a false discovery rate (FDR) $\leq$ 0.05 and an absolute value of $\log_2$ fold change $\geq$ 1 between mutant and wild type (WT) were defined as DEGs. The raw sequence data were deposited in the NCBI Sequence Read Archive (PRJNA749083).

Metabolite Profiling

Metabolites were extracted from petals (500 mg dry weight) with three biological replicates and were analyzed using LC-ESI-MS/MS system at the Metware Biotechnology Co., Ltd. (Wuhan, China). Flavonoids were extracted using the same method at the National Engineering Research Center of Rapeseed (Huazhong Agricultural University, Wuhan, China). The sample extraction and metabolic analysis were explicitly done as previously described (Zhai et al., 2020).

Measurement of Carotenoid and Chlorophyll

For carotenoids analysis, the petals were sampled from double mutants (CRTISO-94-15-5-2, CRTISO-48-8-13-1, CRTISO-7-2-2-4), heterozygous mutant (CRTISO-43-3-4-1, CRTISO-48-8-2-3, CRTISO-48-8-20-5) and WT. Carotenoid pigments extraction and analysis were performed using LC-MS/MS system, as previously described (Lee, 2001; Saladié et al., 2014). Carotenoids were identified based on retention times and absorption spectra as compared to standards. Peak areas were recorded at 286, 348, 473, and 450 nm for phytoene, phytofluene, lycopene, and others, respectively (Xu et al., 2006). The carotenoid levels were quantified using calibration curves prepared with appropriate standards. At least three independent extractions were conducted per sample.

For chlorophyll analysis, the leaves was extracted from double mutants (CRTISO-94-15-5-2, CRTISO-48-8-13-1, CRTISO-7-2-2-4) and WT and quantified using a spectrophotometer, as described previously (Becker, 1994). At least three independent extractions were conducted per sample.

Subcellular Localization

The cDNA sequences of BnaA09.CRTISO and BnaC08.CRTISO without the termination codon were amplified from J9707 using primers CRTISO-15/16 and CRTISO-18/19, respectively (**Supplementary Table 1**). The amplified cDNA fragments were independently cloned into the pMDC83 vector between the *Pac*I and the *Asc*I site, to generate a C-terminal fusion with GFP under control of the cauliflower mosaic virus 35S promoter. The fused construct was introduced into *Nicotiana benthamiana* plants by transient *Agrobacterium* transformation. Samples were observed with a Leica TCSST2 confocal laser microscope (Nikon D760, Tokyo, Japan).

RESULTS

Molecular Cloning and Characterization of *CRTISO* Homologs in *B. napus*

Previous studies revealed that the *CRTISO* gene is essential for regulating carotenoid content and is highly conserved in many plants (Galpaz et al., 2013; Su et al., 2015; Li et al., 2020;

[1]http://www.hi-tom.net/hi-tom/

[2]http://crispr.hzau.edu.cn/CRISPR2/

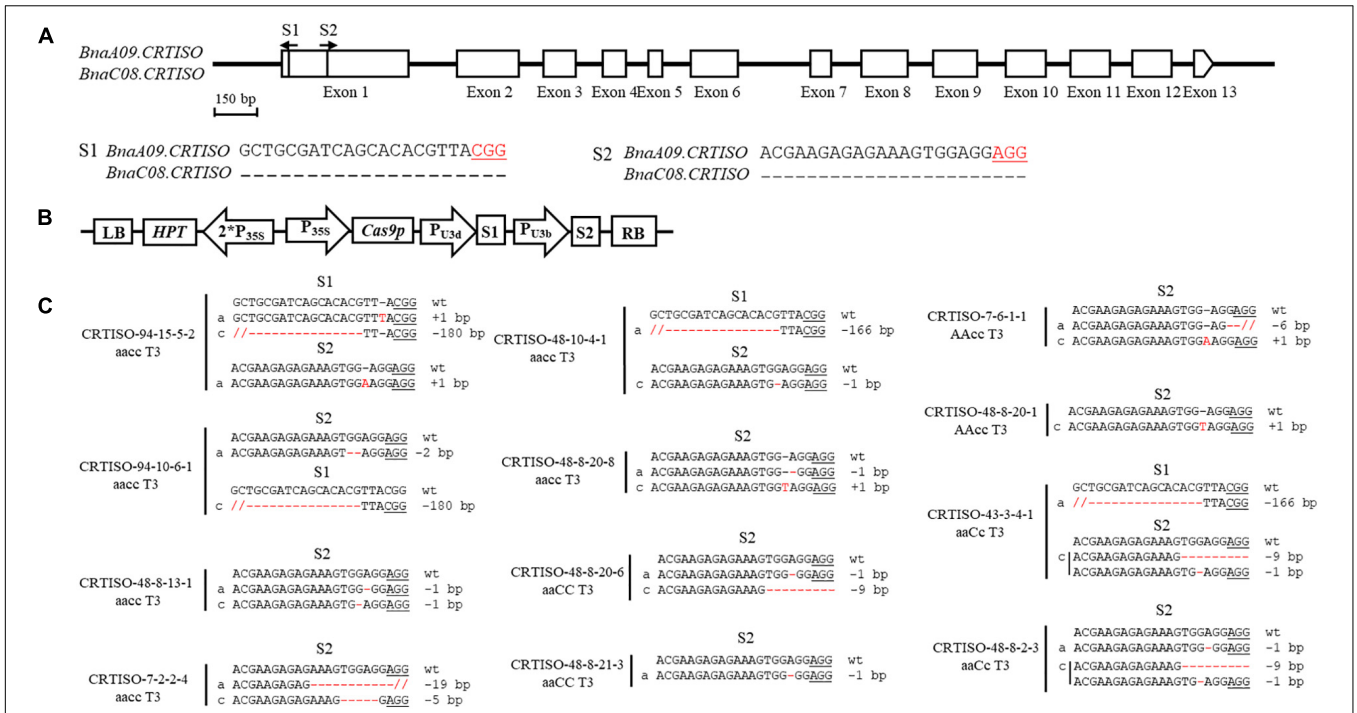


FIGURE 1 | CRISPR/Cas9-induced null mutants of *BnaCRTISO* in *B. napus*. **(A)** The *BnaCRTISO* gene model includes thirteen exons (box) separated by twelve introns (represented by the solid line). The vertical line in the gene model indicates the target site, and the arrow indicates the sgRNA direction. The target sequences are shown with the protospacer adjacent motif (PAM) underlined. **(B)** The CRISPR/Cas9 construct houses the following: a hygromycin resistance cassette consisting of the hygromycin phosphotransferase coding sequence driven by the cauliflower mosaic virus 35S promoter; a Cas9 expression cassette comprising the sequence encoding Cas9 driven by 35S promoter; and two sgRNAs (S1-S2) driven by the U3d and U3b promoters from *Arabidopsis*. **(C)** Sequences at the sgRNA target sites of *BnaCRTISO* homozygous mutants in the T_3 generation. The PAM is underlined, and nucleotide indels are marked in red, with details labeled at right, "a" and "c" represent the mutated alleles of the target gene on *BnaA09.CRTISO* and *BnaC08.CRTISO*, respectively. "*aaCC*," "*AAcc*," and "*aacc*" represent homozygous mutations of the target gene in *BnaA09.CRTISO*, *BnaC08.CRTISO* and both copies, respectively.

Sun et al., 2020). The carotenoid content of a flower directly determines the color of the petals (Su et al., 2015; Zhang et al., 2015; Liu et al., 2020; Zhao et al., 2021). Thus, *CRTISO* is one of the ideal candidates for generating different flower colors in rapeseed. There are two *CRTISO* copies (*BnaA09g49740D* and *BnaC08g44970D*, designated as *BnaA09.CRTISO* and *BnaC08.CRTISO*, respectively) in *B. napus*. In order to check for putative mutations in these target genes, we confirmed their genomic DNA and cDNA sequences in the *B. napus* pure line J9707. Compared to *Arabidopsis* AtCRTISO, the predicted amino acid sequences of both BnaCRTISO contain an important conserved domain, "the amine oxidase" (**Supplementary Figures 1, 2**). which meanings *BnaA09.CRTISO* and *BnaC08.CRTISO* can produce functional amine oxidase proteins.

BnaA09.CRTISO and *BnaC08.CRTISO* was 87.94% identical at the nucleotide level and shared 98.64% amino acid identity, suggesting that both genes encode enzymes with similar functions. The sequence alignment of both *BnaCRTISO* gene copies revealed that polymorphisms distinguished their origins (**Supplementary Figure 3**).

Phylogenetic analysis indicated that all *Brassica* genes were clustered together with *AtCRTISO* and *CRTISO* homologs in other species were clustered on another branch except in tomato (*SolCRTISO*), suggesting that the *CRTISO* gene is differentiated among different species (**Supplementary Figure 4**). *BnaA09.CRTISO* and *BnaC08.CRTISO* were closely related to their homologs in *B. rapa* and *B. oleracea*, respectively (**Supplementary Figure 4**), which is in line with their origination from two diploid progenitors.

Expression Analysis of the *BnaCRTISO* Gene

Analysis of mRNA accumulation patterns of both *BnaCRTISO* copies based on the public RNA-seq data in yellow-flower rapeseed line showed their expression profiles (**Supplementary Figure 5A**). In all cases, *BnaA09.CRTISO* and *BnaC08.CRTISO* had the highest expression levels in the bud, followed by leaf, and the expression levels of both copies in various tissues were comparable. Thus, we further confirmed that the *B. napus* genome contains two functional *AtCRTISO* homologs, *BnaA09.CRTISO* and *BnaC08.CRTISO*.

The expression level of both *BnaCRTISO* copies in J9707 was further examined using qRT-PCR with RNA samples extracted from the petal (**Supplementary Figure 5B**). Consistent with the public RNA-seq data, *BnaC08.CRTISO* had a significantly higher expression level than *BnaA09.CRTISO* in the J9707 petal (**Supplementary Figure 5A**).

Creation of CRISPR/Cas9-Targeted Mutations in *BnaCRTISO*

To generate CRISPR/Cas9-induced knockout mutations in the functional copies of *BnaCRTISO*, two sgRNAs were designed using the CRISPR-P program (Lei et al., 2014). The two sgRNAs [sgRNA1 (S1) and sgRNA2 (S2)] were designed to target the amine oxidase domain, which will induce mutations in the functional domain of the *BnaCRTISO* gene resulting in the formation of non-functional protein (**Figure 1A** and **Supplementary Figure 3**). The designed sgRNAs matched well with both *BnaA09.CRTISO* and *BnaC08.CRTISO* copies (**Figure 1A**). A CRISPR/Cas9 construct containing these two sgRNAs with Cas9 driven by the cauliflower mosaic virus 35S promoter (**Figure 1B**) was produced based on the CRISPR/Cas9 multiplex genome-editing vector previously described by Yang et al. (2017). The resulting construct was transformed into J9707 using *Agrobacterium*-mediated transformation. A total of 105 seedlings were regenerated, of which 96 were transgenic positive. And 72 targeted mutants were identified by Sanger DNA sequencing of the PCR products encompassing the target sites, with 37 plants showing a visible knockout phenotype (i.e., creamy white flower; **Supplementary Table 2**). The overall editing efficiency of the T_0 generations is 75.00%, of which the editing efficiency of S1 and S2 were 16.39 and 73.22%, respectively.

To generate stable lines with targeted mutations, 72 independent T_0 editing lines of *BnaCRTISO* were self-pollinated to produce T_1, T_2, and T_3 progenies. The targeted mutations of progenies from these T_0 lines were verified by Hi-TOM sequencing analysis of the target sites (**Table 1**). The results proved that the mutant genotypes could be stably transmitted to the subsequent generations. A total of 10 T_3 plants with homozygous mutations in *BnaCRTISO* were detected, including two *BnaA09.CRTISO* single mutants, two *BnaC08.CRTISO* single mutants and six *BnaCRTISO* double mutants (**Table 1** and **Figure 1C**). These homozygous mutations at the target sites within *BnaCRTISO* were predicted to cause frameshifts resulting in the production of non-functional proteins (**Supplementary Figure 6**). As expected, all the double mutants could produce the creamy white flower phenotype (**Figure 2B**).

The single mutants showed a similar yellow flower phenotype as WT (**Figures 2A,D,E**), while the heterozygous mutants (*aaCc*) showed a lighter yellow flower phenotype (**Figure 2C**). Thus, both copies of the *BnaCRTISO* gene function redundantly in regulating flower color. In addition, the inner leaves of the double mutants were also more yellowish than those of the WT, single mutants, and heterozygous mutants (**Figures 2F–L**), indicating that the *BnaCRTISO* gene also functions in the leaves.

Off-Target Activity of CRISPR/Cas9 in T_0 and T_1 Transgenic *B. napus* Plants

To ascertain whether off-targeting occurred in the present study, we searched the *B. napus* genome for putative off-target sites with high homology to S1 and S2 according to the CRISPR-P program (Lei et al., 2014). A total of seven putative off-target sites were identified for both sgRNAs (**Supplementary Table 10**), and no off-target editing was detected in T_0 and T_1 by gene-specific primers amplification and Sanger sequencing (**Supplementary Table 10**). This result shows that the off-target effect is negligible when the specificity of each sgRNA is fully considered based on the genome sequence. Thus, the CRISPR/Cas9 system has a high specificity for targeted mutagenesis in *B. napus*.

Subcellular Localization of *BnaCRTISO*

In order to explore the subcellular localization of *BnaA09.CRTISO* and *BnaC08.CRTISO*, GFP was fused to the C terminal of each gene and transiently expressed in tobacco leaves. The green fluorescence signal overlaps closely with the chloroplast red autofluorescence signal observed by confocal microscopy (**Figure 3**). Thus, BnaCRTISO was predicted to be a chloroplast-localized protein, which was consistent with previous reports that carotenoids is synthesized and stored in plastids (Lange and Ghassemian, 2005).

BnaCRTISO Regulates the Expression of Carotenes and Xanthophylls-Related Genes

The petals from flowers during anthesis were collected to compare the expression profiles between *BnaCRTISO* double mutants (CRTISO-94-15-5-2, CRTISO-48-8-13-1, CRTISO-7-2-2-4) and corresponding WT to investigate the transcriptome changes underlying the petal color variation (**Supplementary Table 3**). In subsequent analysis, a total of 41,186 genes were expressed in the developing petals during the same period. In the double mutants and WT, the pearson correlation coefficient between any two of the three biological repeats is very high (R = 0.93–0.98), indicating that the transcriptome sequencing data used in this study is highly reliable (**Supplementary Figure 7**).

Comparing the transcript abundance in these petals, it was observed that there were 2,058 DEGs between the *BnaCRTISO* double mutants and its corresponding WT at the same period (**Supplementary Table 4**). Overall, 953 genes were up-regulated in the double mutant petals, while 1,105 genes were down-regulated, which may be related to the changes observed in petal color (**Supplementary Figure 8**). The GO and KEGG enrichment analysis of these identified DEGs showed that the metabolic processes of carotenoid and flavonoid were significantly enriched among the down-regulated DEGs in mutants relative to WT (**Supplementary Figures 9–11** and **Supplementary Tables 5–7**).

As the targeted mutated gene, *BnaC08.CRTISO* and *BnaA09.CRTISO* were down-regulated by almost eight and three times, respectively. The expression levels of most carotenoid biosynthesis-related genes were significantly up-regulated, such as *BnaPDS3*, *BnaZDS*, *BnaCYP97A3*, *BnaBCH1*, *BnaBCH1*, *BnaZEP*; whereas, five copies of *BnaPSY* were significantly down-regulated. *BnaPSY*, *BnaPDS3*, *BnaZDS*, and *BnaCRTISO* are involved in carotene synthesis; *BnaCYP97A3*, *BnaBCH1* *BnaBCH1*, and *BnaZEP* are related to xanthophylls synthesis. *PSY*, encoding a phytoene synthase, acts like a faucet and plays a significant role at the initial stage in the carotenoid biosynthesis pathway. Moreover, some genes in the carotenoid degradation

TABLE 1 | Genotypic analysis of *BnaCRTISO* mutants and their transmission to T_1, T_2, and T_3 generations.

Plant ID	Generation	Genotype at targets of *BnaCRTISO.A09*		Genotype at targets of *BnaCRTISO.C08*		Petal color
		S1	S2	S1	S2	
CRTISO-94-10	T1	WT	Homo (−2 bp)	Homo (−180 bp)		Creamy white
CRTISO-94-10-6	T2	WT	Homo (−2 bp)	Homo (−180 bp)		Creamy white
CRTISO-94-10-6-1	T3	WT	Homo (−2 bp)	Homo (−180 bp)		Creamy white
CRTISO-94-11	T1	WT	Biallelic	Homo (−180 bp)		Creamy white
CRTISO-94-12	T1	Hetero	Biallelic	Homo (−5 bp)	Homo (−1 bp)	Creamy white
CRTISO-94-13	T1	Hetero	Biallelic	Homo (−180 bp)		Creamy white
CRTISO-94-15	T1	Homo (+1 bp)	Homo (+1 bp)	Homo (−180 bp)		Creamy white
CRTISO-94-15-5	T2	Homo (+1 bp)	Homo (+ 1 bp)	Homo (−180 bp)		Creamy white
CRTISO-94-15-5-2	T3	Homo (+1 bp)	Homo (+1 bp)	Homo (−180 bp)		Creamy white
CRTISO-48-8	T1	Biallelic		WT	Biallelic (−9 bp, −1 bp)	Light yellow
CRTISO-48-9	T1	Homo (−166 bp)		Biallelic		Creamy white
CRTISO-48-10	T1	Homo (−166 bp)		WT	Homo (−9 bp)	Yellow
CRTISO-48-8-13	T2	WT	Homo (−1 bp)	WT	Homo (−1 bp)	Creamy white
CRTISO-48-8-13-1	T3	WT	Homo (−1 bp)	WT	Homo (−1 bp)	Creamy white
CRTISO-48-8-14	T2	WT	Hetero	Homo (−137 bp)		Yellow
CRTISO-48-8-15	T2	WT	Homo (−1 bp)	WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-48-8-18	T2	WT	Homo (−1 bp)	WT	Homo (−9 bp)	Yellow
CRTISO-48-10-4	T2	Homo (−166 bp)		WT	Homo (−1 bp)	Creamy white
CRTISO-48-10-4-1	T3	Homo (−166 bp)		WT	Homo (−1 bp)	Creamy white
CRTISO-48-8-20	T2	WT	Hetero	WT	Chimeric	Yellow
CRTISO-48-8-20-1	T3	WT	WT	WT	Homo (+ 1 bp)	Yellow
CRTISO-48-8-20-5	T3	WT	Homo (−1 bp)	WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-48-8-20-6	T3	WT	Homo (−1 bp)	WT	Homo (−9 bp)	Yellow
CRTISO-48-8-20-8	T3	WT	Homo (−1 bp)	WT	Homo (+ 1 bp)	Creamy white
CRTISO-48-8-21	T2	WT	Hetero	WT	Chimeric	Yellow
CRTISO-48-8-21-3	T3	WT	Homo (−1 bp)	WT	WT	Yellow
CRTISO-48-8-21-7	T3	WT	Hetero	WT	Homo (−9 bp)	Yellow
CRTISO-48-8-2	T2	WT	Homo (−1 bp)	WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-48-8-2-3	T3	WT	Homo (−1 bp)	WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-7-9	T1	Biallelic		Chimeric		Yellow
CRTISO-7-10	T1	Biallelic		WT	Homo (−5 bp)	Yellow
CRTISO-7-12	T1	WT	Homo (−6 bp)	WT	Homo (−5 bp)	Creamy white
CRTISO-7-18	T1	Biallelic		Homo (−180 bp)		Creamy white
CRTISO-7-2	T2	Chimeric		Biallelic		Yellow
CRTISO-7-2-2	T2	WT	Biallelic	WT	Homo (−5 bp)	Yellow
CRTISO-7-2-2-4	T3	WT	Homo (−19 bp)	WT	Homo (−5 bp)	Creamy white
CRTISO-7-2-2-7	T3	WT	Homo (−6 bp)	WT	Homo (−5 bp)	Yellow
CRTISO-7-6	T1	WT	Homo (−6 bp)	WT	Homo (+ 1 bp)	Yellow
CRTISO-7-6-1	T2	WT	Homo (−6 bp)	WT	Homo (+ 1 bp)	Yellow
CRTISO-7-6-1-1	T3	WT	Homo (−6 bp)	WT	Homo (+1 bp)	Yellow
CRTISO-43-2	T1	WT	Homo (−1 bp)	WT	Homo (−1 bp)	Creamy white
CRTISO-43-5	T1	Homo (−166 bp)		WT	Homo (−1 bp)	Creamy white
CRTISO-43-7	T1	WT	Homo (−1 bp)	WT	Homo (−9 bp)	Yellow
CRTISO-43-3	T1	Homo (−166 bp)		WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-43-3-4	T2	Homo (−166 bp)		WT	Biallelic (−9, −1 bp)	Light yellow
CRTISO-43-3-4-1	T3	Homo (−166 bp)		WT	Biallelic (−9, −1 bp)	Light yellow

Hetero, heterozygous; Homo, homozygous; WT, wild type. "–" and "+" indicate the deletion and insertion of the indicated number of nucleotides or nucleotides, respectively; All other targets are wild type except the indicated target.

pathway were also significantly down-regulated, including *BnaCCD4*, *BnaCCD8*, *BnaNCED2*, and *BnaNCED3* (**Figure 4** and **Supplementary Table 7**).

To verify the reliability of the RNA-seq data, 29 DEGs in the petals were selected for qRT-PCR verification analysis. These DEGs include 13 genes involved in the carotenoid

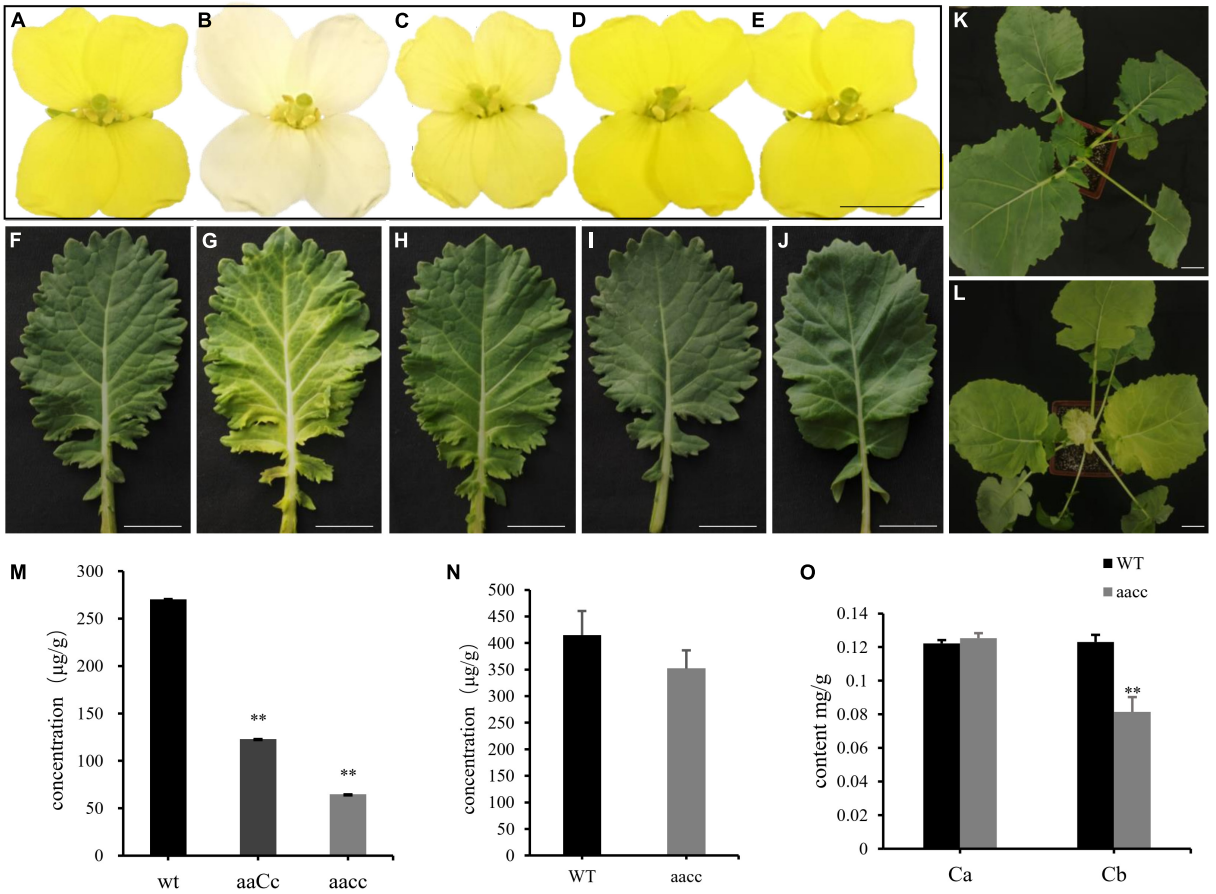


FIGURE 2 | Mutations in the *BnaCRTISO* gene affected the color of petals and leaves. **(A–J)** The phenotype of flower (At the first day after flowering) and leaf in WT, *aacc*, *aaCc*, *AAcc*, *aaCC,* respectively. "*aaCC*," "*AAcc,*" and "*aacc*" represent mutations of the target gene in *BnaA09.CRTISO*, *BnaC08.CRTISO* and both copies, respectively. **(K)** WT seeding. **(L)** *aacc* seeding. Bars, 1 cm. **(M)** The total carotenoid contents (μg/g fresh weight) of petals in aacc, aaCc, WT, respectively. **(N)** The total carotenoid contents of leaves in *aacc*, WT, respectively. **(O)**The content of chlorophyll *a*, *b* in the leaves of WT and *aacc*, respectively. **(M–O)** Data are presented as means ± SE ($n \geq 3$); *t*-test was used for statistical analysis between the mutant and its WT (**$P > 0.01$).

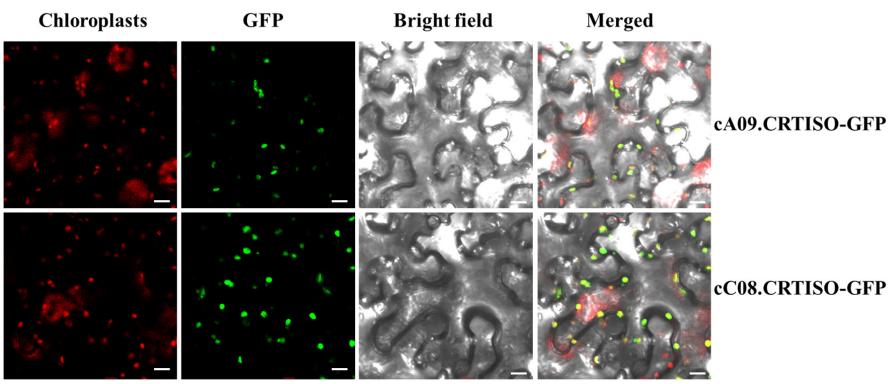


FIGURE 3 | Subcellular localization of BnaA09.CRTISO and BnaC08.CRTISO in tobacco leaves. The fusion proteins generated using the coding sequence (cA09.CRTISO-GFP and cC08.CRTISO-GFP) were independently introduced into tobacco leaves. Scale bar = 10 μm.

synthesis, five genes involved in the flavonoid metabolism, and 11 randomly selected genes. Linear regression analysis showed that the correlation coefficient between the transcript levels assessed by the two analytic systems was very high ($R = 0.812$; **Supplementary Figure 12**), which further confirmed the reliability of the RNA-seq data.

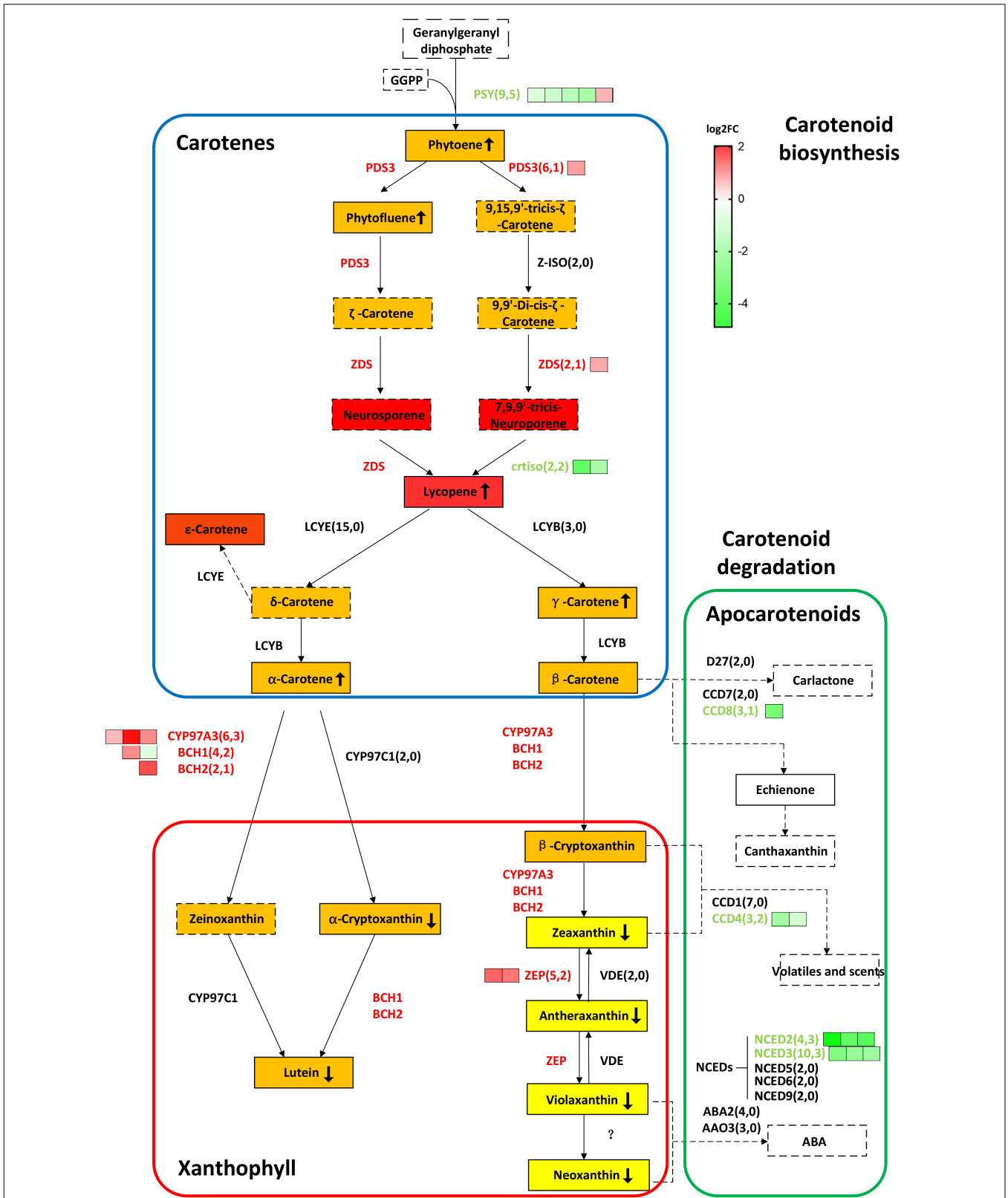


FIGURE 4 | *BnaCRTISO* regulates genes related to carotenes and xanthophyll in rapeseed. In this study, the gene copy numbers from the rapeseed genome and DEGs are listed in parentheses. The $\log_2$ fold change (*crtiso*/WT) of DEGs of different copies of the same gene is represented by green (down-regulated) or red

(Continued)

FIGURE 4 | (up-regulated) squares arranged from left to right. Gene names are represented in capital letters, and corresponding mutants are represented in small letters. Genes that are up-regulated and down-regulated are represented in red and blue, respectively. The background color of the box indicates the color of the substance. The detected carotenoids in present study were indicated in the solid box, and undetected carotenoids in present study were indicated in the dotted box. The up or down arrow in the box indicates increase or decrease of the corresponding substance content, respectively. PSY, PHYTOENE SYNTHASE; PDS3, 15-*cis*-zeta-carotene isomerase; Z-ISO, 15-*cis*-zeta-carotene isomerase; ZDS, zeta-carotene desaturase; CRTISO, carotenoid isomerase; LCYE, lycopene beta/epsilon cyclase protein; LCYB, lycopene cyclase; BCH1, beta-hydroxylase 1; BCH2, beta-carotene hydroxylase 2; CYP97A3, cytochrome P450, family 97, subfamily A, polypeptide 3; CYP97C1, cytochrome P450 superfamily protein; ZEP, zeaxanthin epoxidase; VDE, non-photochemical quenching 1; NCED5,nine-*cis*-epoxycarotenoid dioxygenase 5; NCED3, nine-*cis*-epoxycarotenoid dioxygenase 3; NCED6, nine-*cis*-epoxycarotenoid dioxygenase 6; NCED9, nine-*cis*-epoxycarotenoid dioxygenase 9; NCED2, nine-*cis*-epoxycarotenoid dioxygenase 2; CCD4, nine-*cis*-epoxycarotenoid dioxygenase 4; ABA2, NAD(P)-binding Rossmann-fold superfamily protein; AAO3, abscisic aldehyde oxidase 3; CYP707A2, cytochrome P450, family 707, subfamily A, polypeptide 2; CYP707A4, cytochrome P450, family 707, subfamily A, polypeptide 4; CYP707A3, cytochrome P450, family 707, subfamily A, polypeptide 3; CYP707A1, cytochrome P450, family 707, subfamily A, polypeptide 1; D27, beta-carotene isomerase D27-like protein; CCD7, carotenoid cleavage dioxygenase 7; CCD8, carotenoid cleavage dioxygenase 8.

Together, these results indicated the importance of the *BnaCRTISO* gene in the metabolic pathway of carotenoid. This also further illustrates the complex regulatory mechanisms of the *BnaCRTISO* gene in carotenoid synthesis of rapeseed.

Targeted Mutations in *BnaCRTISO* Change Pigment Concentrations

To assess the impact of targeted mutation of *BnaCRTISO* on the carotenoid metabolic pathway, double (*aacc*: CRTISO-94-15-5-2, CRTISO-48-8-13-1, CRTISO-7-2-2-4) and heterozygous (*aaCc*: CRTISO-43-3-4-1, CRTISO-48-8-2-3, CRTISO-48-8-20-5) mutant T_3 lines were grown in the field with their WT control. The total carotenoid content of petal and leaf samples were analyzed using spectrophotometer. It showed that the total carotenoid content of petals were significantly decreased in *aacc* mutants compared with *aaCc* and WT (*aacc* < *aaCc* < WT) (**Figure 2M**). The similar trend was observed in the total carotenoid content of leaves (**Figure 2N**). The two types of chlorophyll (chlorophyll *a* and chlorophyll *b*) concentration were further measured in leaves. And the chlorophyll *b* concentration of *aacc* mutants were significantly decreased relative to WT; whereas, no difference in chlorophyll *a* concentrations were observed between *aacc* mutants and WT (**Figure 2O**). This result indicated that *BnaCRTISO* may be involved in the anabolism of chlorophyll *b* in leaves.

Furthermore, carotenoid were analyzed using an LC-MS/MS system. Almost all carotene and xanthophyll contents showed a significant difference among *aacc*, *aaCc,* and WT. The carotene content of *aacc* was significantly increased when compared to the *aaCc* mutant and WT (*aacc* > *aaCc* > WT); while the xanthophyll content was significantly decreased (*aacc* < *aaCc* < WT) (**Figure 5** and **Supplementary Table 8**). (E/Z)-phytoene, phytofluene, and lycopene are the main components of carotene. In this study, no lycopene was detected in the petals of WT; unexpectedly, the lycopene contents of the *aacc* and *aaCc* mutant were significantly increased (*aacc* > *aaCc* > WT; **Supplementary Table 8**). It should be noted that lycopene produced by the BnaCRTISO enzyme catalyzes prolycopene conversion in the carotenoid pathway. This result indicates the occurrence of another pathway for lycopene biosynthesis or degradation in the *BnaCRTISO* mutant, which converts prolycopene into all-*trans* lycopene through a non-enzymatic reaction under light illumination (Tuan et al., 2011; Su et al., 2015). In the xanthophyll metabolic pathway, the contents of α-cryptoxanthin (synthetic lutein precursor substance), β-cryptoxanthin (zeaxanthin precursor substance), lutein, zeaxanthin, antheraxanthin, violaxanthin, and neoxanthin were all significantly decreased in *aacc* mutants (**Figure 5** and **Supplementary Table 8**). These changes of xanthophylls content agreed well with the phenotypic variations in these materials: the higher the xanthophylls content, the yellower the petals.

Targeted Mutations in *BnaCRTISO* Change Flavonoid Metabolites in Petals

The flavonoid metabolic profiling of petal samples from the double-mutant (CRTISO-94-15-5-2, CRTISO-48-8-13-1, CRTISO-7-2-2-4) and its WT was analyzed using LC-ESI-MS/MS system. Most of the identified flavonoid metabolites show a significant difference between the double-mutant and WT petals (**Supplementary Table 8**). Among them, naringenin chalcone, the main flavonoid metabolites that constitutes yellow petals, were significantly decreased in the mutant. It is consitent with the transcriptomic analysis of the mutant petals, which indicates that the expression of naringenin chalcone metabolic genes (*BnaC4H*, *BnaTT4*, *BnaTT7*, *BnaPAL2*, *Bna4CL3*) were decreased in the double mutant (**Figure 6**, **Supplementary Figure 12**, and **Supplementary Table 9**). The content of other anthocyanins (apigenin and luteolin) in the mutants also decreased significantly (**Figure 6** and **Supplementary Table 9**). Thus, these results are in line with the phenotypes and the transcriptomic analysis of the mutant petal.

DISCUSSION

The CRISPR/Cas9-Targeted Mutation in *BnaCRTISO* Is a Promising Strategy to Change Flower Color in Rapeseed

Although flower color in *B. napus* is a very important trait and has attracted much interest since 1929 (Pearson, 1929), the underlying molecular mechanisms have not been well characterized. The white flower is an ornamental trait and ideal phenotypic marker in assisted breeding. Studies has also indicated that the white flower trait of *B. napus* is closely linked with high erucic acid (Liu et al., 2004). Therefore, it is significant to develop phenotypic markers and improve the

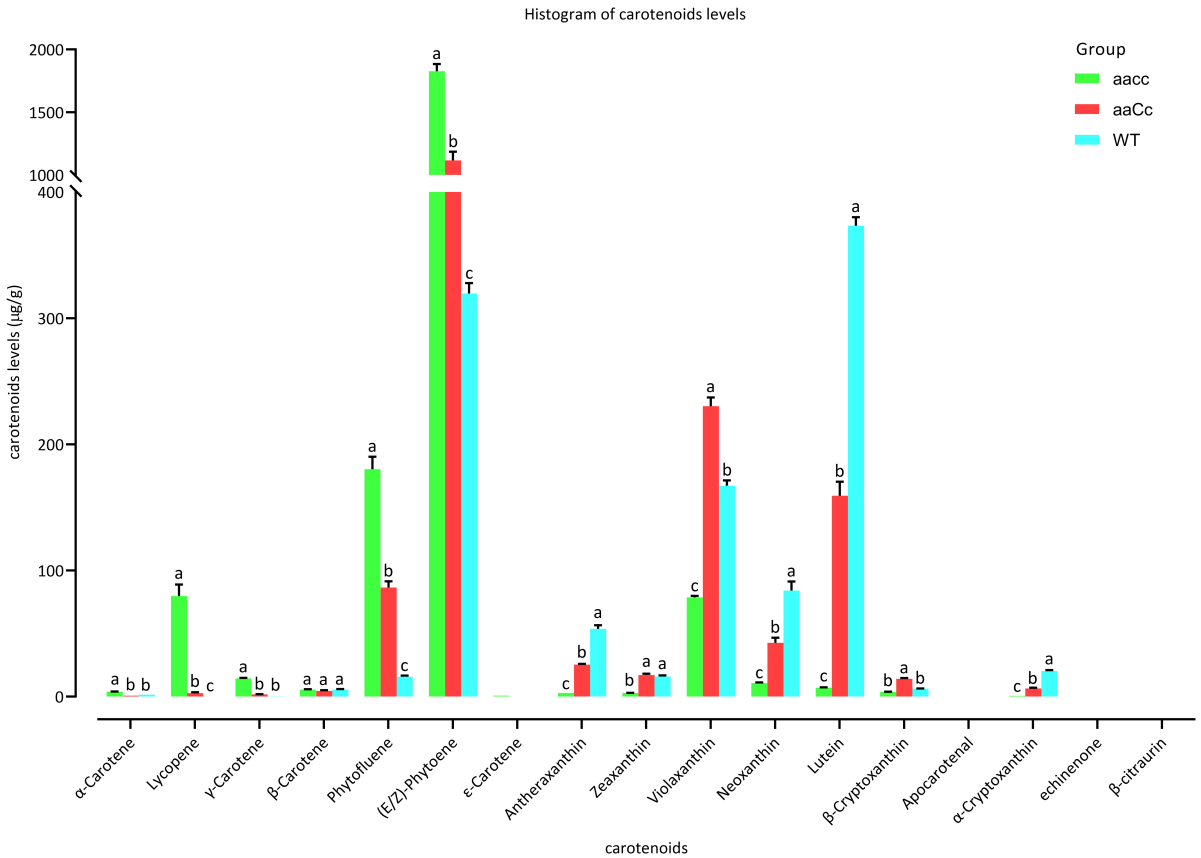


FIGURE 5 | The different carotenoids levels in the petals of mutants and WT. *aacc*, homozygous double mutants; *aaCc*, the heterozygous mutants with a lighter yellow flower phenotype; WT, wild type; LSD-test was used for statistical analysis between the aacc and WT (Lower case letter indicates the significant difference at 0.05 level).

ornamental value of low erucic acid rapeseed. To this end, an effective approach is needed to produce targeted mutations in these well-conserved *CRTISO* homologs in *B. napus*. The newly developed CRISPR/Cas9 technology provides a powerful approach for creating novel allelic variations. Thus far, it has been successfully utilized to modify several important agronomic traits (multilocular silique, plant height, architecture, pod shatters resistance, etc.) in rapeseed by generating specific gene knockout (Yang et al., 2017, 2018; Li C. et al., 2018; Zhai et al., 2019).

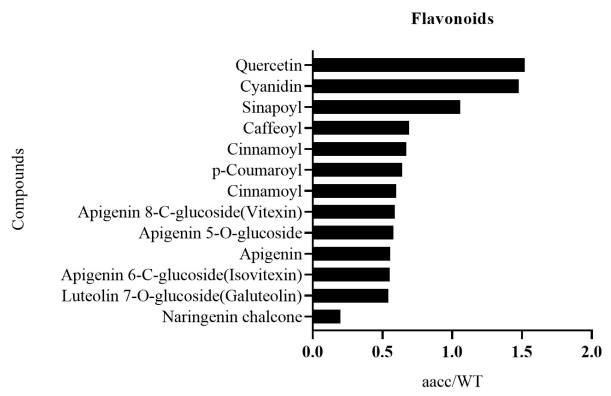


FIGURE 6 | The identified flavonoid metabolites in petals with significant difference between the *BnaCRTISO* double mutant (*aacc*) and WT. The level of most of these metabolites decreased in mutant petals. And they are listed in the order of decreased fold changes (*aacc*/WT) from top to down.

This study shows the successful utilization of CRISPR/Cas9 for targeted mutations of the *BnaCRTISO* gene in rapeseed with an editing efficiency of 75.00%, which is higher than those reported in previous research (14.4–45.0%) (Yang et al., 2017; Zhai et al., 2019; Ahmar et al., 2021). The editing efficiency at the S2 target site (73.22%) is higher than S1 (16.39%), indicating that the efficiency of the sgRNA promoter U3b is higher than that of U3d as our previous reports (Yang et al., 2017; Zhai et al., 2019; Ahmar et al., 2021). In this study, the targeted mutations of the two functional copies of *BnaCRTISO* in rapeseed generated the creamy white petal phenotype (**Supplementary Table 2** and **Figure 2**), indicating that the *BnaCRTISO* gene is essential for flower color regulation and highly conservative in *Brassica* plants. The phenotypic analysis shows that the flower color of the double mutants is creamy white, while the *BnaA09.CRTISO* and *BnaC08.CRTISO* single mutants showed a comparable phenotype to that of the WT. Hence, both copies of the *BnaCRTISO* gene

have functional redundancy in regulating flower color formation, and their contribution is equal. In addition, the heterozygous mutant (*aaCc*) identified in this study had a biallelic mutation (−9, −1 bp; e.g., CRTISO-8-15) on the *BnaC08.CRTISO* with "aa" genotype, and the plant shows a slight change in petal color (**Figure 2B**). While the double allelic mutation (−6, −19 bp; e.g., CRTISO-7-2-2) on *BnaA09.CRTISO* with "cc" genotype shows no change in petal color (**Table 1**). A homozygous −9 bp mutation on *BnaC08.CRTISO* with "aa" genotype (CRTISO-48-10, CRTISO-48-8-18, CRTISO-48-8-20-6, CRTISO-48-8-21-7, CRTISO-43-7) also shows no change in petal color (**Table 1**). Thus, these results indicated that different mutation types have different effects on BnaCRTISO enzyme activity.

BnaCRTISO Plays an Important Role in the Accumulation of Pigments in Petals

Flavonoids and carotenoids are the pivotal pigments for the formation of most flower colors (Zhang et al., 2015). It has been reported that *BnaC3.CCD4* and *CmCCD4a* contribute to the white color formation in the petals of *B. napus* and *chrysanthemum*, respectively, by degrading carotenoids into colorless compounds (Ohmiya et al., 2006; Zhang et al., 2015). Recently, Zhao et al. (2021) identified a mutation in a *PDS3* gene in carotenoid biosynthesis pathway causes yellowish-white petals in rapeseed (Zhao et al., 2021). To date, no gene regulating the white flower traits other than the *CCD4* and *PDS3* gene has been reported in rapeseed.

In *Arabidopsis*, the *crtiso* mutant exhibiting partial inhibition of lutein biosynthesis in light-grown tissue and the accumulation of poly-*cis*-carotene precursors in dark-grown tissue with cotyledons colors changing from yellow to orange (Hyoungshin, 2002). The loss of function of the *BrCRTISO* gene confers orange color to the inner leaves and induces changes in flower color from yellow to orange in Chinese cabbage (*B. rapa*) (Su et al., 2015). Recently, the targeted editing of *BoaCRTISO* changed the leaf color from green to yellow, with the significant reduction of the concentrations of carotenoids and chlorophylls in Chinese kale (Sun et al., 2020). The *crtiso* mutant in tomato accumulates prolycopene instead of all-*trans*-lycopene, which could result in the production of orange fruit (Kato et al., 2007). It shows that the *CRTISO* gene in different plants can effectively change the carotenoid content in the flower, leaf, or fruit. In rapeseed, the cotyledon color of the *BnaCRTISO* double mutant showed a similar color change as that in *B. rapa* under the dark treatment (**Supplementary Figure 13**). Additionally, the petal color of the *BnaCRTISO* mutant changed from yellow to creamy white, which is different from that of *B. rapa* and *Arabidopsis*. It implies that the function of *CRTISO* in allotetraploid *B. napus* has become more complex in the course of evolution, and more information about this gene involved in the carotenoid pathway is yet to be ascertained.

PSY is the first key enzyme in the synthesis pathway, and its expression is regulated by feedback from upstream and downstream genes and metabolites (Yonekura-Sakakibara et al., 2019). The *BnaPSY* gene is also an important gene encoding the rate-limiting enzyme in the carotenoid synthesis pathway (Zhou et al., 2015). Down-regulation of the *BnaPSY* gene leads to a decrease in the total carotenoids content in the petals of the *BnaCRTISO* double mutants. Moreover, the xanthophyll content also decreased drastically in the *BnaCRTISO* mutants, and there are almost no xanthophylls in the petals of the double mutants. In contrast, the content of lycopene synthesized by the catalysis of BnaCRTISO protein increased significantly in the *BnaCRTISO* mutant. This result is consistent with the changes in the carotenoids content measured in the leaves of *Arabidopsis CRTISO* mutants (Isaacson, 2002). The expression of key genes *BnaPDS3* and *BnaZDS* in the carotenoid synthesis pathway was also significantly up-regulated. This shows that the steady down-regulation of *BnaCRTISO* resulted in the up-regulation of *BnaPDS3* and *BnaZDS*, which promote the accumulation of prolycopene and the mass synthesis of lycopene *via* a non-enzymatic pathway (**Figure 4**).

The expression levels of several xanthophyll synthetic-related genes (*BnaCYP97A3*, *BnaBCH1*, *BnaBCH2*, *BnaZEP*) in the mutants are up-regulated. Compared with the slight change in the synthetic substrate content, the content of the downstream substances (α-cryptoxanthin, lutein, zeaxanthin, antheraxanthin, violaxanthin, neoxanthin) are decreased significantly (**Figure 4** and **Supplementary Table 8**). Interestingly, the expression of genes that inhibit ABA hormone signal transduction, such as *BnaHAI1*, *BnaABI1*, and *BnaHAI2*, are significantly down-regulated (**Supplementary Table 7**). In contrast, the expression of the ABA hormone signaling receptor *BnaPYL4* is significantly up-regulated (**Supplementary Table 7**). It shows that the ABA hormone signal transduction process is accelerated in the mutant. Previoly reports indicated that ABA is synthesized from xanthophylls, e.g., zeaxanthin, violaxanthin and neoxanthin (Li and Walton, 1990; Parry et al., 1990). Thus, the low content of xanthophylls in the *BnaCRTISO* mutant is probably due to its degradation.

In the past, researchers created various types of light-colored flowers in different plants by changing the expression of the *CHS* gene to generate either white or light white flowers (Wang et al., 2018). In this study, the expression of multiple copies of the *BnaCHS* gene, which is the upstream gene regulating the flavonoid metabolism, was significantly down-regulated in the *BnaCRTISO* mutant resulting in the closure of the entire flavonoid synthesis pathway and a sharp decrease in the flavonoids content in petal (**Supplementary Table 7**). *BnaF3′H* is a key gene regulating for anthocyanin synthesis in the flavonoid synthesis pathway (Dubos et al., 2010; He et al., 2013), and its expression in the mutant is also greatly down-regulated. Collectively, these results indicate that the *BnaCRTISO* is an important gene involved in both the carotenoid and flavonoid pathways.

At present, there is no direct correlation between the carotenoid pathway and flavonoid synthesis pathway in plants (Grotewold, 2006; Cazzonelli and Pogson, 2010). It is known that they jointly regulate the color of flowers through the accumulation and mixing of pigments to produce their colored substances. In this study, a creamy white flower mutant was obtained by the mutation of the *BnaCRTISO* gene. Through metabolome and transcriptome analysis, we observed that

the loss of *BnaCRTISO* gene function affects not only the expression of related genes in the carotenoid pathway, but also the expression of key genes involved in the flavonoid synthesis pathway. This reveals that some unknown mechanism of interactions exists between the carotenoid and flavonoid pathway in *B. napus* that are worthy of further study.

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

YZ and CF conceived the study and designed the experiments. LH performed the experiments. KY performed the bioinformatic analysis and wrote the manuscript. YY, ML, MG, and SD helped in the material sampling. LH, OA, and CF helped in the revision of this manuscript. CF supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

The study was financially supported by the National Natural Science Foundation of China (31671279; 31971976; 31371240), and Fundamental Research Funds for the Central Universities (2662015PY172).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.801456/full#supplementary-material

REFERENCES

Ahmar, S., Zhai, Y., Huang, H., Yu, K., Hafeez, U. K. M., Shahid, M., et al. (2021). Development of mutants with varying flowering times by targeted editing of multiple SVP gene copies in *Brassica napus* L. *Crop J.* doi: 10.1016/j.cj.2021.03.023

Becker, EW (ed.) (1994). *Microalgae Biotechnology and Microbiology*. Cambridge: Cambridge University Press, 56–62.

Braatz, J., Harloff, H. J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174, 935. doi: 10.1104/pp.17.00426

Breitenbach, J., and Sandmann, G. (2005). zeta-Carotene cis isomers as products and substrates in the plant poly-cis carotenoid biosynthetic pathway to lycopene. *Planta* 220, 785–793. doi: 10.1007/s00425-004-1395-2

Cazzonelli, C. I. (2011). Carotenoids in nature: insights from plants and beyond. *Funct. Plant Biol.* 38, 833–847. doi: 10.1071/FP11192

Cazzonelli, C. I., and Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trend Plant Sci.* 15, 266–274. doi: 10.1016/j.tplants.2010.02.003

Cunningham, F. X., and Gantt, E. (1998). Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 557–583. doi: 10.1146/annurev.arplant.49.1.557

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). Myb transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581.

Fraser, P. D., and Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* 43, 228–265. doi: 10.1016/j.plipres.2003.10.002

Galpaz, N., Burger, Y., Lavee, T., Tzuri, G., Sherman, A., Melamed, T., et al. (2013). Genetic and chemical characterization of an EMS induced mutation in *Cucumis melo* CRTISO gene. *Arch. Biochem. Biophys.* 539, 117–125. doi: 10.1016/j.abb.2013.08.006

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57, 761–780. doi: 10.1146/annurev.arplant.57.032905.105248

Han, F., Cui, H., Zhang, B., Liu, X., Yang, L., Zhuang, M., et al. (2019). Map-based cloning and characterization of BoCCD4, a gene responsible for white/yellow petal color in B. *oleracea*. *BMC Genomics* 20:242. doi: 10.1186/s12864-019-5596-2

He, H., Ke, H., Keting, H., Qiaoyan, X., and Silan, D. (2013). Flower colour modification of chrysanthemum by suppression of F3′H and overexpression of the exogenous *Senecio cruentus* F3′5′H gene. *PLoS One* 8:e74395. doi: 10.1371/journal.pone.0074395

Holt, N. E., Zigmantas, D., Valkunas, L., Li, X. P., Niyogi, K. K., and Fleming, G. R. (2005). Carotenoid cation formation and the regulation of photosynthetic light harvesting. *Science* 307, 433–436. doi: 10.1126/science.1105833

Hu, L., Zhang, H., Yang, Q., Meng, Q., Han, S., Nwafor, C. C., et al. (2018). Promoter variations in a homeobox gene, BnA10.LMI1, determine lobed leaves in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 131, 1–10. doi: 10.1007/s00122-018-3184-5

Huang, Z., Ban, Y., Bao, R., Zhang, X., Xu, A., and Ding, J. (2014). Inheritance and gene mapping of the white flower in *Brassica napus* L. *N. Z. J. Crop Hortic. Sci.* 42, 111–117.

Hyoungshin, P. (2002). Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. *Plant Cell* 14, 321–332. doi: 10.1105/tpc.010302

Isaacson, T. (2002). Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. *Plant Cell* 14, 333–342. doi: 10.1105/tpc.010303

Kato, M., Matsumoto, H., Ikoma, Y., Kuniga, T., Nakajima, N., and Yoshida, T. (2007). Accumulation of carotenoids and expression of carotenoid biosynthetic genes and carotenoid cleavage dioxygenase genes during fruit maturation in the juice sacs of 'Tamami,' 'Kiyomi' Tangor, and 'Wilking' Mandarin. *Engei Gakkai Zasshi* 76, 103–111. doi: 10.2503/jjshs.76.103

Lange, B. M., and Ghassemian, M. (2005). Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry* 66, 413–451. doi: 10.1016/j.phytochem.2004.12.020

Lee, H. S. (2001). Characterization of carotenoids in juice of red navel orange (Cara Cara). *J. Agric. Food Chem.* 49, 2563–2568. doi: 10.1021/jf001313g

Lei, Y., Lu, L., Liu, H., Li, S., Xing, F., and Chen, L. (2014). CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* 7, 1494–1496. doi: 10.1093/mp/ssu044

Li, C., Hao, M., Wang, W., Wang, H., Fan, C., and Chu, W. (2018). An efficient CRISPR/Cas9 platform for rapidly generating simultaneous mutagenesis of multiple gene homoeologs in allotetraploid oilseed rape. *Front. Plant Sci.* 9:442. doi: 10.3389/fpls.2018.00442

Li, C., Ji, J., Wang, G., Li, Z., Wang, Y., and Fan, Y. (2020). Over-Expression of LcPDS, LcZDS, and LcCRTISO, genes from wolfberry for carotenoid biosynthesis, enhanced carotenoid accumulation, and salt tolerance in tobacco. *Front. Plant Sci.* 11:119. doi: 10.3389/fpls.2020.00119

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., et al. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front. Plant Sci.* 9:559. doi: 10.3389/fpls.2018.00559

Li, Y., and Walton, D. (1990). Violaxanthin is an abscisic acid precursor in water-stressed dark-grown bean leaves. *Plant Physiol.* 92, 551–559. doi: 10.1104/pp.92.3.551

Liu, X., Tu, J., Chen, B., and Fu, T. (2004). Identification of the linkage relationship between the flower colour and the content of erucic acid in the resynthesized *Brassica napus* L. *Acta Genet. Sin.* 31, 357–362. doi: 10.1046/j.1439-0523.2002.00711.x

Liu, Y., Ye, S., Yuan, G., Ma, X., Heng, S., Yi, B., et al. (2020). Gene silencing of BnaA09.ZEP and BnaC09.ZEP confers orange color in *Brassica napus* flowers. *Plant J.* 104, 932–949. doi: 10.1111/tpj.14970

Lv, P., Li, N., Liu, H., Gu, H., and Zhao, W. (2015). Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars. *Food Chem.* 174, 52–59. doi: 10.1016/j.foodchem.2014.11.022

Ma, X., Zhang, Q., and Zhu, Q. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *ScienceDirect Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007

Matsufuji, H., Ishikawa, K., Nunomura, O., Chino, M., and Taketa, M. (2007). Anti-oxidant content of different coloured sweet peppers, white, green, yellow, orange and red (*Capsicum annuum* L.). *Int. J. Food Sci. Technol.* 42, 1482–1488.

Montserrat, R. B., Inés, G., Carlos, V., Maria, B., and Rosa, M. E. (2015). Carotenoids from haloarchaea and their potential in biotechnology. *Mar. Drugs* 13, 5508–5532. doi: 10.3390/md13095508

Nisar, N., Li, L., Shan, L., Khin, N. C., and Pogson, B. (2015). Carotenoid metabolism in plants. *Mol. Plant* 8, 68–82.

Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., and Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in Chrysanthemum petals. *Plant Physiol.* 142, 1193–1201. doi: 10.1104/pp.106.087130

Parry, A., Babiano, M., and Horgan, R. (1990). The role of cis-carotenoids in abscisic acid biosynthesis. *Planta* 182, 118–128. doi: 10.1007/BF00239993

Patrick, S., Marta, R. F., Ian, C. C., Florian, W., and Ralf, W. (2018). Establishment of an *Arabidopsis* callus system to study the interrelations of biosynthesis, degradation and accumulation of carotenoids. *PLoS One* 13:e0192158. doi: 10.1371/journal.pone.0192158

Pearson, O. (1929). A dominant white flower color in *Brassica oleracea* L. *Am. Nat.* 63, 561–565.

Quazi, M. H. (1988). Interspecific hybrids between *Brassica napus* L. and B. *oleracea* L. developed by embryo culture. *Theor. Appl. Genet.* 75, 309–318.

Sakai T., Iwabuchi M., Kohno-Murase J., Liu H. J., and Imamura J., (1995). "Transfer of radish CMS-restorer gene into *Brassica napus* by intergeneric protoplast fusion," in *Proceedings of the 9th International Rapeseed Congress*, Cambridge, 3–5.

Saladié, M., Wright, L. P., Garcia-Mas, J., Rodriguez-Concepcion, M., and Phillips, M. A. (2014). The 2-C-methylerythritol 4-phosphate pathway in melon is regulated by specialized isoforms for the first and last steps. *J. Exp. Bot.* 65, 5077–5092. doi: 10.1093/jxb/eru275

Shahid, M., Cai, G., Zu, F., Zhao, Q., Qasim, M., Hong, Y., et al. (2019). Comparative transcriptome analysis of developing seeds and silique wall reveals dynamic transcription networks for effective oil production in *Brassica napus* L. *Int. J. Mol. Sci.* 20:1982. doi: 10.3390/ijms20081982

Su, T., Yu, S., Zhang, J., Yu, Y., Zhang, D., and Zhao, X. (2015). Loss of function of the carotenoid isomerase gene BrCRTISO confers orange color to the inner leaves of Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Mol. Biol. Rep.* 33, 648–659. doi: 10.1007/s11105-014-0779-0

Sun, B., Jiang, M., Zheng, H., Jian, Y., and Tang, H. (2020). Color-related chlorophyll and carotenoid concentrations of Chinese kale can be altered through CRISPR/Cas9 targeted editing of the carotenoid isomerase gene BoaCRTISO. *Hortic. Res.* 7:161. doi: 10.1038/s41438-020-00379-w

Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., and Li, L. (2018). Carotenoid metabolism in plants: the role of plastids. *Mol. Plant* 11, 58–74. doi: 10.1016/j.molp.2017.09.010

Tuan, P. A., Kim, J. K., Park, N. I., Lee, S. Y., and Park, S. U. (2011). Carotenoid content and expression of phytoene synthase and phytoene desaturase genes in bitter melon (*Momordica charantia*). *Food Chem.* 126, 1686–1692. doi: 10.1016/j.foodchem.2010.12.058

Walter, M. H., and Strack, D. (2011). Carotenoids and their cleavage products: biosynthesis and functions. *Nat. Prod. Rep.* 28, 663–692. doi: 10.1039/c0np00036a

Wang, Z., Yu, Q., Shen, W., El Mohtar, M., Zhao, X., and Gmitter, F. G. (2018). Functional study of CHS gene family members in citrus revealed a novel CHS gene affecting the production of flavonoids. *BMC Plant Biol.* 18:189. doi: 10.1186/s12870-018-1418-y

Wurtzel, V. E. T. (2009). Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiol.* 150, 562–572. doi: 10.1104/pp.109.137042

Xu, C.-J., Fraser, P. D., Wang, W.-J., and Bramley, P. M. (2006). Differences in the carotenoid content of ordinary citrus and lycopeneaccumulating mutants. *J. Agric. Food Chem.* 54, 5474–5481. doi: 10.1021/jf060702t

Yang, H., Wu, J., Tang, T., Liu, K., and Dai, C. (2017). CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. *Sci. Rep.* 7:4877. doi: 10.1038/s41598-017-07871-9

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, M. H. U., et al. (2018). Precise editing of CLAVATA genes in *Brassica napus* L. regulates multilocular silique development. *Plant Biotechnol. J.* 16, 1322–1335. doi: 10.1111/pbi.12872

Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* 10:943. doi: 10.3389/fpls.2019.00943

Zhai, Y., Cai, S., Hu, L., Yang, Y., Amoo, O., and Fan, C. (2019). CRISPR/Cas9-mediated genome editing reveals differences in the contribution of indehiscent homologues to pod shatter resistance in *Brassica napus* L. *Theor. Appl. Genet.* 132, 2111–2123. doi: 10.1007/s00122-019-03341-0

Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., et al. (2020). 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, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., et al. (2015). Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in *Brassica* species. *New Phytol.* 206, 1513–1526. doi: 10.1111/nph.13335

Zhang, B., Lu, C. M., Kakihara, F., and Kato, M. (2010). Effect of genome composition and cytoplasm on petal colour in resynthesized amphidiploids and sesquidiploids derived from crosses between *Brassica rapa* and *Brassica oleracea*. *Plant Breed.* 121, 297–300. doi: 10.1046/j.1439-0523.2002.722295.x

Zhao, C., Safdar, L. B., Xie, M., Shi, M., Liu, S., Dong, Z., et al. (2021). Mutation of the phytoene desaturase 3 gene causes yellowish-white petals in *Brassica napus*. *Crop J.* 9, 1124–1134.

Zhou, X., Welsch, R., Yang, Y., Riediger, M., and Yuan, H. (2015). *Arabidopsis* OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3558–3563. doi: 10.1073/pnas.1420831112

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: 14 July 2022
doi: 10.3389/fpls.2022.940765

BnaA03.ANS Identified by Metabolomics and RNA-seq Partly Played Irreplaceable Role in Pigmentation of Red Rapeseed (*Brassica napus*) Petal

Pengfei Hao[1†], *Han Liu*[1,2†], *Baogang Lin*[1], *Yun Ren*[3], *Lan Huang*[1], *Lixi Jiang*[4] and *Shuijin Hua*[1*]

[1] Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2] Yongding Agriculture and Rural Bureau of Longyan, Longyan, China, [3] Huzhou Agricultural Science and Technology Development Center/Huzhou Academy of Agricultural Sciences, Huzhou, China, [4] College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Sarvajeet Singh Gill,
Maharshi Dayanand University, India

Reviewed by:
Zhansheng Li,
Insititute of Vegetables and Flowers (CAAS), China
Cunmin Qu,
Southwest University, China

***Correspondence:**
Shuijin Hua
sjhua1@163.com

[†]These authors have contributed equally to this work

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

Received: 10 May 2022
Accepted: 17 June 2022
Published: 14 July 2022

Citation:
Hao P, Liu H, Lin B, Ren Y, Huang L, Jiang L and Hua S (2022) BnaA03.ANS Identified by Metabolomics and RNA-seq Partly Played Irreplaceable Role in Pigmentation of Red Rapeseed (Brassica napus) Petal. Front. Plant Sci. 13:940765. doi: 10.3389/fpls.2022.940765

Colorful flowers of rapeseed (*Brassica napus* L.) have been a hotspot for researchers, but the underlying mechanisms of pigment formation still need to be clarified. In this study, two stages of unopened rapeseed petals with red, white, and yellow colors were selected to identify the metabolites and genes involved in red pigment formation. Metabolomic analysis showed that flavonoids enriched the most co-differentially accumulated metabolites among all categories, and showed higher accumulation in red petal rapeseed than in white and yellow petal ones. RNA-seq analysis showed that among co-differentially expressed genes involved in red pigment formation, genes involved in anthocyanin (belonging to flavonoids) biosynthesis pathway were largely regulated by *ANS*, *DFR*, and *UF3GT*. The expression of those genes was higher in red petals of rapeseed than in white and yellow petals ones as well. Results of RNA interference of *BnaA03.ANS* in red rapeseed altered petal colors from raspberry red to beige red and zinc yellow under different interference levels, with the contents of pelargonidin, cyanidin, lutein, neoxanthin, β-carotene, and lycopene significantly decreased. However, overexpression of *BnaA03.ANS* in yellow rapeseed petals did not change the color of yellow petals. This study confirmed the important function of flavonoids, especially anthocyanins on red pigment formation, and for the first time, identified the irreplaceable role of *BnaA03.ANS* on red-flowered rapeseed.

Keywords: carotenoids, metabolomics, overexpression, RNA interference, RNA-seq

INTRODUCTION

Brassica napus L. (AACC, $2n = 38$), an allopolyploid species derived from the hybridization of *Brassica rapa* (AA, $2n = 20$) and *Brassica oleracea* (CC, $2n = 18$), is an important cash crop not only showing edible and industrial oil purpose but also exhibiting great ornamental value (Chalhoub et al., 2014; Fu et al., 2016). For example, Hubei Province, China, owns over 25.68 million hectares of rapeseed, the tourism of which caused by rapeseed "flower sea" has contributed to more than half of the county's GDP in 2011 (Fu et al., 2016).

Flower color is critical for ornamental and landscaping utilization by providing aesthetically pleasing scenes. The traditional flower color of rapeseed is yellow and some mutants with light to dark yellow, and with the exploitation of the ornamental values, various petal colors were bred, such as white, orange, pink, red, and purple (Yin et al., 2019). However, the mechanisms for regulating different petal colorations are still unclear.

Carotenoids, flavonoids (namely, anthocyanins, flavones, and flavonols), and betalain biosynthesis are reported as the three most important pathways which produce secondary metabolites that contribute to natural petal color display. For instance, carotenoid and betalain pathways contributed to yellow and red, while orange, yellow, red, and blue were attributed to the anthocyanin pathway (Tanaka et al., 2008). More than that, petal tissue structure, epidermal cell shapes, and pH were also reported to influence the formation of petal colors (Vignolini et al., 2015; Zhao and Tao, 2015). Anthocyanins are the main group of flavonoids, and play a crucial role in plant color development, ranging from pink to blue and purple. Over 100 anthocyanins have been identified, primarily originating from six common types, namely, pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin (Veitch and Grayer, 2008). Previous studies showed that the color differences are highly correlated to different anthocyanin contents and components. By measuring the anthocyanin content of a series of butterfly pea petals, Kazuma et al. (2003) found that the content of anthocyanins was significantly higher in blue petals than in other colors while no anthocyanins were identified in white color petals. As for cineraria, delphinidin and cyanidin mainly contributed to blue and red flower colors, pink flowers were mainly determined by cyanidin and pelargonidin, while delphinidin and cyanidin were the core anthocyanins in purple flowers (Sun et al., 2009). Moreover, Zhang et al. (2011) found that the main pigment deposited in Lagenaria red petals was cyanidins, and pelargonidins primarily leaned toward scarlet.

Based on the irreplaceable role, anthocyanin biosynthesis has risen to be a hotspot of research for plant secondary metabolism and the key genes that participated in its biosynthetic pathway in plants have been clarified (Cheynier et al., 2013). *CHALCONE SYNTHASE* (*CHS*) is the key gene encoding the first enzyme in the anthocyanin biosynthesis pathway, hence influencing the coloration of flower petals. Ectopic expression of *CHS1* gene from Freesia hybrid in petunia altered flower color from white to pink (Sun et al., 2015), and transgenic tobacco plants with *CHS* gene from Malus crabapple showed higher anthocyanin accumulation and a deep red color than the wild-type (Tai et al., 2014). *CHALCONE ISOMERASE* (*CHI*) encodes the second enzyme in anthocyanin biosynthesis and catalyzes the formation of chalcone. The decrease of *CHI* expression level in the petal of asters, tobacco, and carnations was found to lead to a greater accumulation of chalcone and turn the color into yellow (Nishihara et al., 2005). *DIHYDROFLAVONOL 4-REDUCTASE* (*DFR*) is another key gene encoding an enzyme that transfers three types of dihydroflavonols to their corresponding colorless anthocyanins with NADPH. Zhao et al. (2012) found that in different herbaceous peony organs, the highest expression level of *DFR* was observed in petals which accumulated the most anthocyanins, and similar results were also reported in Asian lily and gentian, suggesting the important role of *DFR* in flower color formation (Nakatsuka et al., 2003, 2005). *ANTHOCYANIDIN SYNTHASE* (*ANS*) is a key gene in the late stage of anthocyanin biosynthesis, which catalyzes the leucoanthocyanin to colored anthocyanidin (Heller et al., 1985). It was identified that *ANS* is a small gene family and these genes have been successfully cloned from several ornamental plants, such as *Forsythia supensa*, herbaceous peony, and gerbera (Rosati et al., 1999; Wellmann et al., 2006; Zhao et al., 2012). Rosati et al. (1999) found that null expression of *ANS* in *Forsythia suspensa* resulted in little anthocyanins accumulation in petals, and the absence of expression of *ANS* was found to be the underlying reason for the color change in lisianthus flowers (Shimizu et al., 2011). To date, the cloning and functional verification of genes that control petal color have paved for gene engineering of plants to obtain colorful exhibitions. For example, in *Brassica napus*, Liu et al. (2020) reported that silencing of *BnaA09.Zep* and *BnaC09.Zep* through Crispr/Cas9 confers orange color in *B. napus* petals. By ectopic overexpression of the *OvPAP2* (*Orychophragmus violaceus*) gene, Fu et al. (2018) successfully produced red anthers and petals in yellow oilseed rape. Zhang et al. (2015) reported that disruption of *BnaC03.CCD4* gene enhanced the accumulation of carotenoids leading to the transfer of petal color from white to yellow.

However, how the anthocyanin pathway regulates rapeseed petal color formation and the detailed molecular mechanisms remain unclear, and few studies have reported color-related genes or elaborated on the molecular regulation mechanism underlying anthocyanin-based variation in oilseed rape petal colors (Sagawa et al., 2016; Nikolov, 2019). The function of *ANS*, the key genes in anthocyanin biosynthesis, on flower coloration in *Brassica napus* is not reported in comparison with others. Here, we performed a metabolomics and RNA-seq study on two different stages of unopened petals of red, pure white, and yellow petal rapeseed lines, aiming at elaborating the pigment formation and development mechanisms in *B. napus*. *BnaA03.ANS* was identified as a high expression co-differentially expressed gene (co-DEG) according to RNA-seq and qRT-PCR. Disruption of *BnaA03.ANS* converted rapeseed petal color from raspberry red to beige red or zinc yellow, while overexpression of *BnaA03.ANS* made no change in petal color, but anthocyanins and carotenoids contents showed large differences. This study first comprehensively elaborated on the pigment formation and development mechanisms from genes to metabolites which contributed a lot to the foundation for colorful rapeseed breeding and the function of *BnaA03.ANS* on red color formation was first identified in *B. napus*.

MATERIALS AND METHODS

Plant Materials and Sampling

Three *B. napus* lines with contrasting petal colors were used, which were Zhehuhong (red, abbreviated as V1), Zhehubai (white, abbreviated as V2), and Zheyou 50 (yellow, abbreviated as V3). The transgenic plants with overexpression and RNAi were obtained from Zheyou 50 and Zhehuhong, respectively.

The rapeseed plants including the transgenic lines were planted in a glass solar greenhouse with three replications in a completely randomized block design at the Zhejiang Academy of Agricultural Science, China.

At the flowering stage, full plump unopened buds from the main inflorescence with <5 flowers were selected, then all the petals in each bud were separated out, and the developmental stages of unopened petals were defined as follows: stage 1, petals with 5 mm length and colors close to pale; stage 2, petals with 6–7 mm length; stage 3, petals with 8–9 mm length; and stage 4, petals with 10 mm length and were strongly pigmented. The petals in stage 1 and stage 4 of V1, V2, and V3 were utilized for RNA-seq and metabolomic analysis with 3 biological replicates. Petals at all stages were utilized for stereoscopic imaging, and petals at stage 4 were used for carotenoids and anthocyanins quantification. Petal colors were compared according to the RAL method (Long et al., 2011).

Metabolomic Analysis

Petals at stage 1 and stage 4 of V1, V2, and V3 were prepared, named V1-1, V1-4, V2-1, V2-4, V3-1, and V3-4, respectively, and 3 biological replicates were set. About 50 mg sample was weighed out for supernatant preparation. Agilent 1290 Infinity II UHPLC system coupled to an Agilent 6545 UHD and Accurate-Mass Q-TOF/MS was used for liquid chromatography-mass spectrometry (LC-MS) analysis. The chromatographic column used was Waters XSelect HSS T3 (2.5 μm, 100 mm × 2.1 mm). Raw data were converted to common (mz.data) using Agilent MassHunter Qualitative Analysis B.08.00 software (Agilent Technologies, USA). Then all data went through internal standard normalization and weight normalization. Visualization matrices containing sample name, m/z-RT pair, and peak area were obtained. After editing, the data matrices were imported into SIMCA-P 14.1 (Umetrics, Umea, Sweden), mean-centered and scaled to Pareto variance. Then, a multivariate analysis was conducted. $|\log_2\text{FoldChange}|>1$ and $p \leq 0.05$ were determined as differentially accumulated metabolites (DAMs). DAMs Venn diagram, principal component analysis (PCA), Gene Ontology (GO), and KEGG enrichment (Kyoto Encyclopedia of Genes and Genomes, KEGG) were also conducted for further analysis.

RNA-seq Analysis

The preparation of samples was the same as metabolomic analysis. Total RNA was extracted using a polysaccharide and polyphenol total RNA isolation kit (Bioteke, Beijing, China). HISAT2 was used to align all clean reads against the reference genome *Brassicanapus*.Annotation_v4.1. FPKM values were used to calculate the expression level of genes. $|\log_2\text{FoldChange} >1$ and $p \leq 0.05$ were determined as differentially expressed genes (DEGs). Venn diagram of DEGs, PCA, GO, and KEGG was also conducted for further analysis.

Quantification of Carotenoids and Anthocyanins Contents

Contents of carotenoids and anthocyanins in 9 rapeseed petals (V1-4, V2-4, V3-4, and each containing 3 biological replicates) were determined through HPLC methods. Carotenoids were performed according to Cao et al. (2012) and the contents of each carotenoid were determined as previously described (Morris et al., 2004).

The components of anthocyanins were determined according to Sun et al. (2014) with several modifications. Approximately 100 mg petal samples were weighed out and 800 μl methanol was added and vortex for 1 min, then shock under 4°C for 30 min. A total of 600 μl supernatant was prepared and concentrated to dry, then 200 μl methanol was added to redissolve the samples. The supernatants were used for LC-MS/MS analysis after centrifuging under 12,000 rpm 4°C for 10 min. The MS parameters were: ESI ion source, 35 arb curtain gas, 7 arb collision gas, 4,500 V ion spray voltage, 450°C temperature, 55 arb ion source gas1, and 55 arb ion source gas2. Multiple reaction monitoring was used for parameter acquisition and MultiQuant software was used for data calculation.

Plasmid Construction and Transformation

Petal-specific expression promoter *XY355* and the open reading frame of *BnaA03.ANS* was amplified and cloned into the *Pme1-Pac1* and *Pac1-Asc1* sites of pMDC83, respectively, to construct overexpression plasmid *pXY355::BnaA03.ANS*. The plasmid was introduced into the *Agrobacterium tumefaciens* strain GV3101 and transformed into yellow-flowered rapeseed V3.

The target gene fragments were amplified using *BnaA03.ANS* RNAi primers (**Supplementary Table 1**), and then ligated into the pNC-Cam1304-RNAi vector using the Nimble Cloning kit (Chinese Academy of Tropical Agricultural Sciences) to construct an *ANS* interference vector. The constructed vector was transformed into *Agrobacterium tumefaciens* GV3101, and the genetic transformation of rapeseed hypocotyl mediated by *Agrobacterium* was used to transform it into red rapeseed V1. Both Nimble Cloning kit and pNC-Cam1304-RNAi vector were donated by Mr. Yan Pu (Chinese Academy of Tropical Agricultural Sciences). The genetic transformation methods were referred to by Zhou et al. (2012).

qRT-PCR Analysis

Total RNA was extracted from petal samples using a polysaccharide and polyphenol total RNA isolation kit (Bioteke, Beijing, China) and qRT-PCR was performed according to Xu et al. (2017). All primers used in the study were listed in **Supplementary Table 1**.

Statistics

Data analysis was performed using IBM SPSS v.22.0 statistical software. Duncan's multiple range test was used to evaluate significant treatment effects at the significance level of $p \leq 0.05$. MultiQuant software was used to integrate the curves and calculate the contents of anthocyanins and carotenoids according to each standard curve.

RESULTS

Phenotype Characterization of all *B. napus* Materials

In this experiment, the phenotype of V1, V2, V3, RNAi, and the overexpression lines were recorded (**Figure 1**). The opened petal color of V1, V2, and V3 matched RAL3017

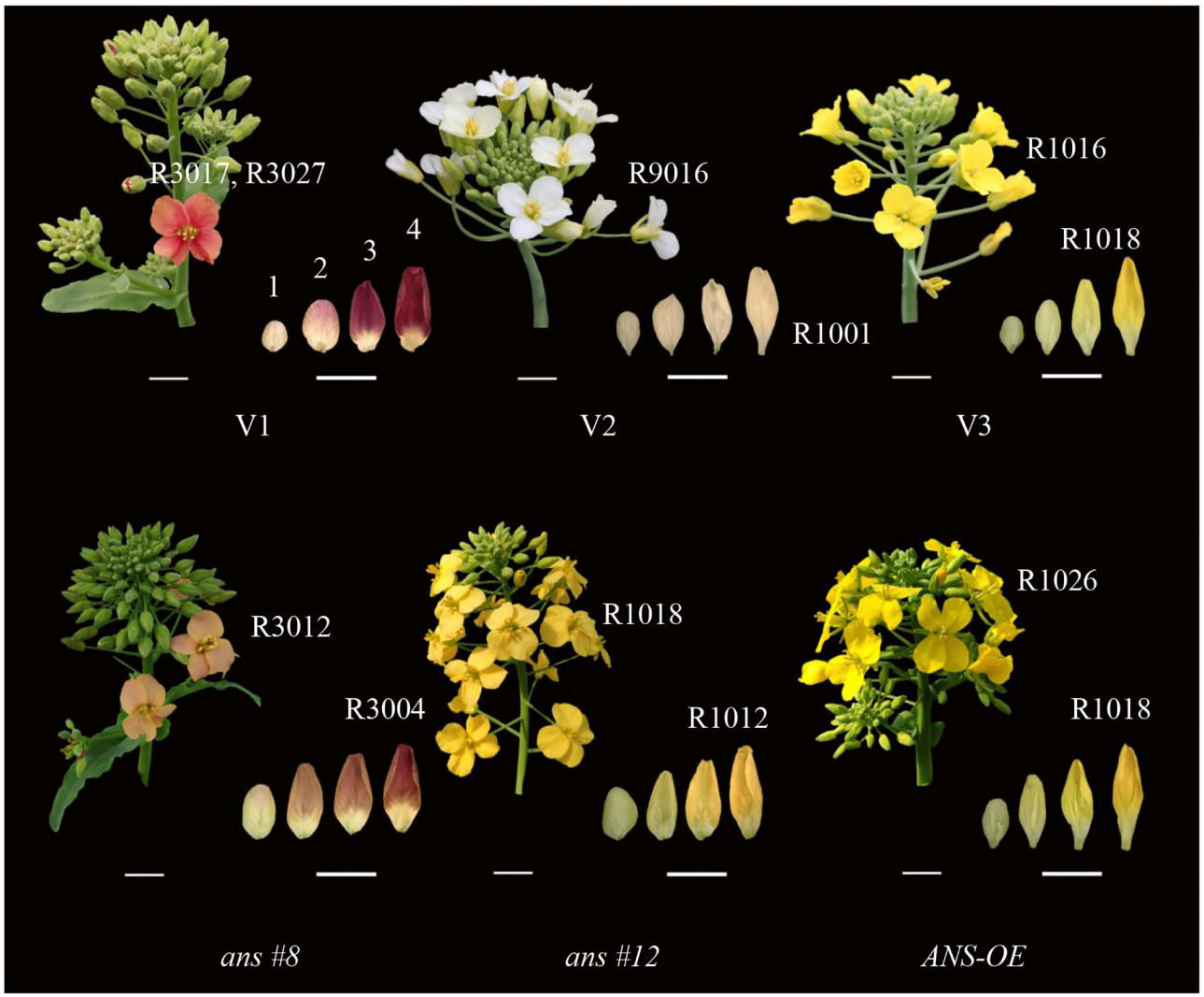


FIGURE 1 | Inflorescence individuals and unopened petals phenotypes of *Brassica napus* with red (Zhehuhong, V1), white (Zhehubai, V2), yellow (Zheyou 50, V3) color, and the transgenic lines (*ans#8* and *ans#12* indicated RNAi-1 and RNAi-2 plants; *ANS-OE indicated* overexpression plants). Numbers on the top of unopened petals in V1 represented 4 development stages. R numbers represent colors alignment to RAL methods. Bars at the left bottom represent 1 cm and bars at the right bottom represent 0.5 cm.

(rose), RAL9016 (traffic white), and RAL1016 (sulfur yellow), respectively. Different from the edge of the V1 petal, the color of the center matched RAL3027 (raspberry red). While the unopened petals were deeper than those opened petals to some extent, the colors of petals at stage 4 matched RAL4004 (claret violet), RAL1001 (beige), and RAL1018 (zinc yellow), respectively, in V1, V2, and V3.

Petal's color became shallow to some degree in *BnaA03.ANS*RNAi lines, with the color matched to RAL3012 (beige red, count 62.5% of all RNAi lines) and RAL1018 (zinc yellow, count 37.5% of all RNAi lines), while the color of unopened petals at T4 stage matched RAL3004 (purple red) and RAL1012 (lemon yellow), respectively. Overexpression of *BnaA03.ANS* did not alter the unopened petal color with the opened petal color match to RAL1026 (luminous yellow).

Identification and Functional Analysis of Differentially Accumulated Metabolites in V1, V2, and V3 at Stage 1 and Stage 4

The result of PCA showed that V1-1 and V1-4 separated significantly from other groups (**Figures 2A,B**). Under positive mode, a total of 163 and 170 co-DAMs were identified between V1 vs. V2 and V1 vs. V3 at stage 1 and stage 4 (Co-DAMs-T1 and Co-DAMs-T4), respectively. Among them, 46+2+103 metabolites were also co-differentially accumulated in stage 1

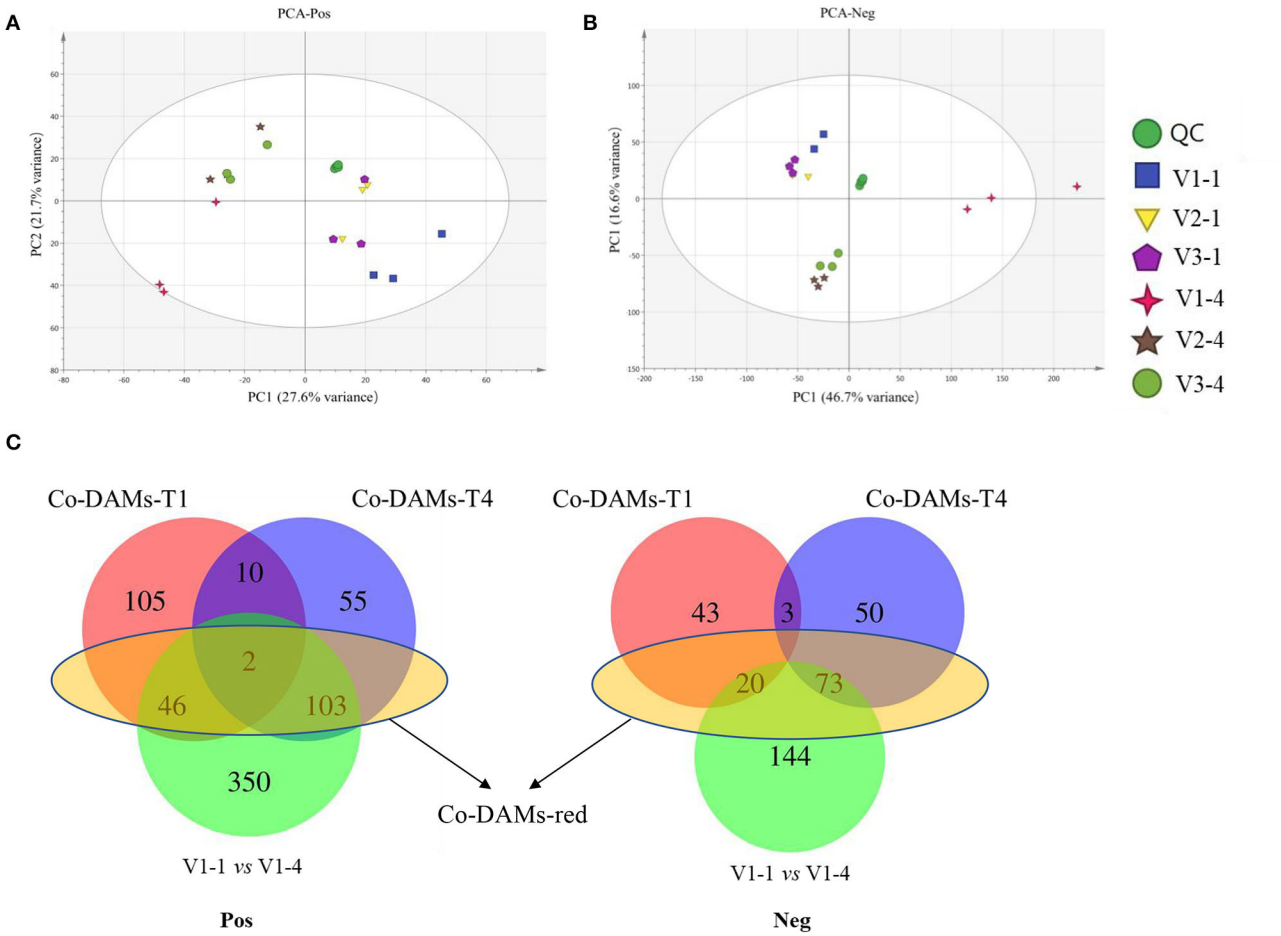


FIGURE 2 | PCA analysis and Venn diagrams of metabolomic data. **(A)** PCA analysis of metabolomic data under positive mode. **(B)** PCA analysis of metabolomic data under negative mode. **(C)** Venn diagram analysis of differentially accumulated metabolites. V1-1, V2-1, and V3-4 represent the unopened petals of V1, V2, and V3 at stage 1. V1-4, V2-4, and V3-4 represent the unopened petals at stage 4. Co-DAMs-T1 represents the common differentially accumulated metabolites between V1 vs. V2 and V1 vs. V3 at stage 1. Co-DAMs-T4 represents the common differentially accumulated metabolites between V1 vs. V2 and V1 vs. V3 at stage 4. Co-DAMs-red represents the metabolites that were considered key metabolites functioning in red color. Pos represents the positive mode and Neg represents the negative mode.

and stage 4 of V1, which were determined as Co-DAMs-red metabolites functioning on the red color formation (**Figure 2C**). The accumulation level of each sample group was shown as a heatmap after normalization processing (**Figure 3A**), and the top 3 classes were carboxylic acids and derivatives (12 DAMs), fatty acyls (11 DAMs), and benzene and substituted derivatives (8 DAMs). Furthermore, KEGG analysis was performed aiming at these 46+2+103 DAMs, and 10 KEGG pathways were enriched significantly (**Figure 3C**).

Under negative mode, a total of 66 and 126 co-DAMs were identified between V1 vs. V2 and V1 vs. V3 at stage 1 and stage 4, respectively. Among them, 20+73 metabolites were also co-differentially accumulated at stage 1 and stage 4 of V1 (**Figure 2C**). The accumulation level of each sample group was shown as a heatmap after normalization processing (**Figure 3B**), with flavonoids (15 DAMs) being the most class. Furthermore, KEGG analysis was performed aiming at these 20+73 DAMs, and 8 KEGG pathways were enriched significantly (**Figure 3D**).

Identification and Functional Analysis of Differentially Expressed Genes in V1, V2, and V3 at Stage 1 and Stage 4

For RNA-seq analysis, raw reads, clean reads, total genes, and sequenced genes of each sample were shown in **Supplementary Table 2**. PCA results showed that these 6 group samples had a distinct separation except for V2 and V3 at stage 4 (**Figure 4A**). A total of 21 genes were selected randomly to validate the reliability of RNA-seq data by qRT-PCR analysis. Results showed that the relative coefficients of the Log_{10}Value of the 21 genes between RNA-seq and qRT-PCR results at group V1-1 vs. V1-4, V2-1 vs. V2-4, and V3-1 vs. V3-4 were 0.7392, 0.7457, and 0.7229, which showed the high reliability of the RNA-seq data (**Figures 4B–D**).

The result of Venn diagrams showed that at stage 1, 1,530 co-DEGs (1,474+56, named Co-DEGs-T1) were identified between V1 vs. V2 and V1 vs. V3, and a total of 9,919 co-DEGs (9,471+178, named Co-DEGs-T4) were identified between V1

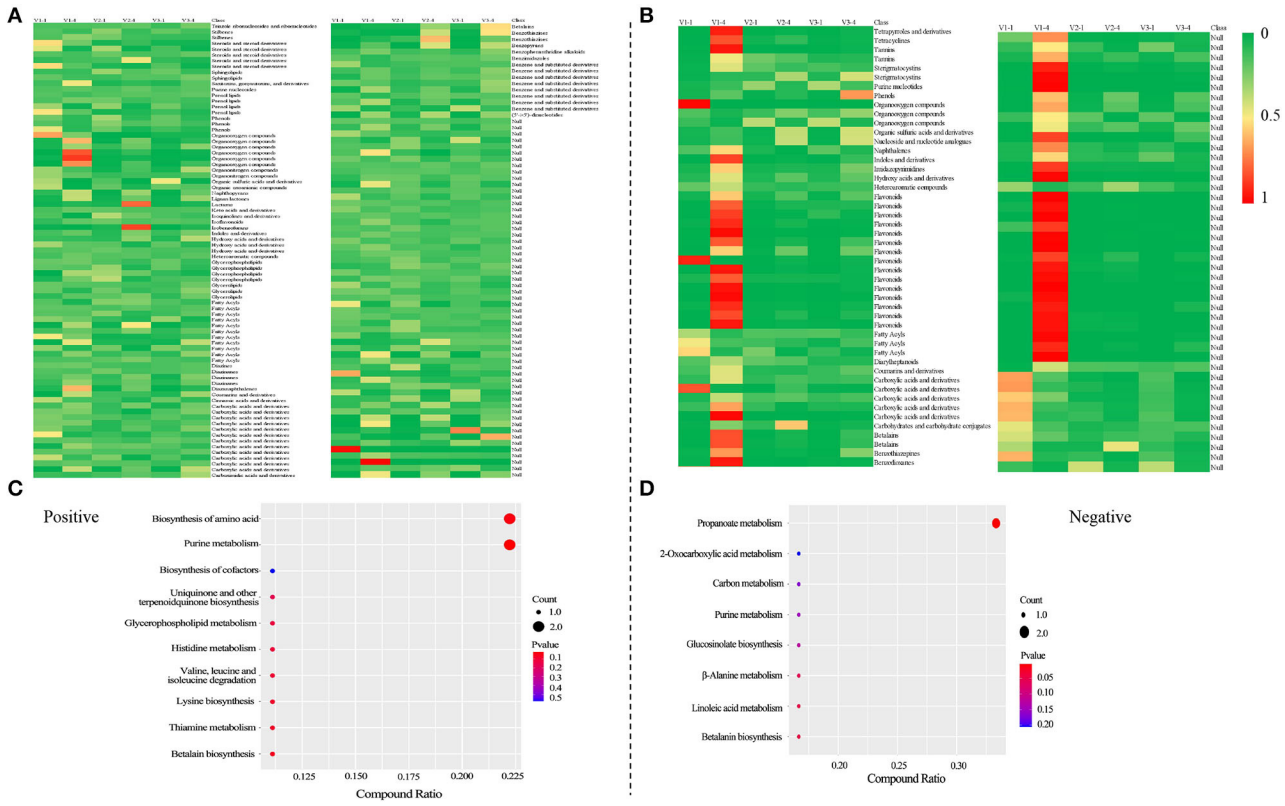


FIGURE 3 | Heatmaps of the accumulation level of differentially accumulated metabolites after normalization and KEGG analysis of Co-DAMs-red. Heatmaps were drawn after normalization, that is take the sum of each metabolite accumulation value in all 6 samples as 1, and calculate the ratio in the sum. **(A,C)** were at positive mode; **(B,D)** were at negative mode.

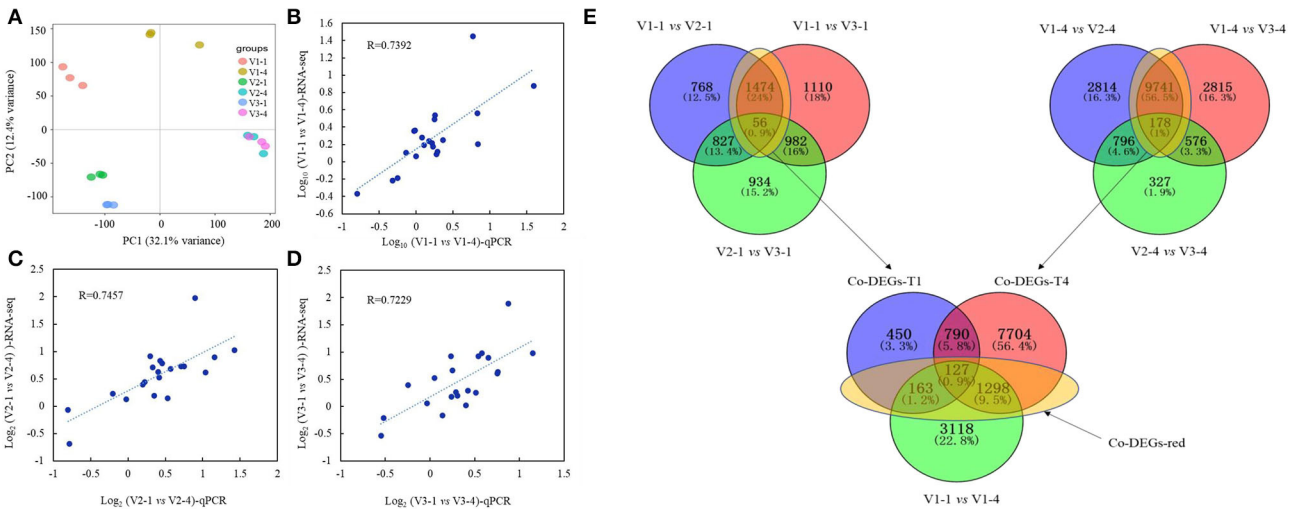


FIGURE 4 | PCA analysis, qRT-PCR validation, and Venn diagrams of RNA-seq data. **(A)** PCA analysis of RNA-seq data; **(B–D)** represent the validation and correlation analysis of RNA-seq and qRT-PCR results of 21 selected genes. **(B)** V1-1 vs. V1-4; C, V2-1 vs. V2-4; D, V3-1 vs. V3-4. **(E)** Venn diagram of common differentially expressed genes (co-DEGs) between V1, V2, and V3 at stage 1 and stage 4. Co-DEGs-T1 represents the genes that are co-differentially expressed in V1 vs. V2 and V1 vs. V3 at stage 1. Co-DEGs-T4 represents the genes that are co-differentially expressed in V1 vs. V2 and V1 vs. V3 at stage 4. Co-DEGs-red represents the genes that were considered key genes that are functioning in red color formation.

vs. V2 and V1 vs. V3 at stage 4. Among them, 163+127+1,298 co-DEGs were also differentially expressed between Co-DEGs-T1 and Co-DEGs-T4 with V1-1 vs. V1-4. These co-DEGs (163+127+1,298, named DEGs-red) were considered to function on the difference between red with white, red with yellow and light red with deep red petal color (**Figure 4E**).

Response of DEGs-Red Participating in Anthocyanin Biosynthesis Pathway, MYB and bHLH Transcription Factors

Aiming at DEGs-red, a total KEGG analysis was conducted and the results were shown in **Supplementary Figure 1**. Anthocyanin biosynthesis pathway was drawn and the heatmap of relative gene expression level was exhibited. Among them, *ANS* was significantly upregulated in five comparison groups (**Figure 5**). In addition, MYB, WD40, and bHLH transcription factor family are reported to be the three main transcription factors that affected petal color formation, and the heatmap of MYB and bHLH were shown in **Supplementary Figure 2**. What's more, qRT-PCR analysis was performed to further verify the expression level of genes relative to petal color (**Supplementary Figure 3**). Among them, *ANS2*, *ANS4*, *PAP2*, *DFR1*, *DFR2*, and *UF3GT* genes had significantly higher expression amounts in red petals than in white and yellow at both stage 1 and stage 4. The expression amount of *CHS*, *CHI*, *F3H*, *F3'H*, and *MYB111* was significantly higher in V2 at stage 1 and was higher in V2 at stage 4 for *PAP2* and *MBY61* while it was higher in V3 at stage 1 for *MYB5* and *MYC1*.

Influence of RNAi and Overexpression of *BnaA03.ANS* on Petal Colors and Pigment Contents

Semiquantitative RT-PCR was also conducted to identify the expression level of 4 copies of *ANS* genes in V1, V2, and V3. The results showed that *BnaA03.ANS* appeared evident difference in all lines, with the brighter band in V1 (**Figure 6A**). Hence, *BnaA03.ANS* was chosen to be the candidate gene for RNAi and overexpression validation.

RNA interference of *BnaA03.ANS* showed that, opened petal colors of V1 turned from rose and raspberry red to zinc yellow, with the expression level of *BnaA03.ANS* decreased by 700-fold −1,400-fold, and turn to beige red when the expression level of *BnaA03.ANS* was only 71.5% or 28.25% of V3 (**Figure 6B**).

Analysis of anthocyanin and carotenoid profiles indicated that pelargonidin and cyanidin were largely decreased by 97.0% and 37.5% in *ans#7*, and 97.8% and 62.1% in *ans#8* compared with V1, respectively. As for carotenoids, lutein, neoxanthin, and β-carotene were also significantly decreased, with 82.5%, 76.2%, and 80.1% and 68.6%, 12.4%, and 60% lower in *ans#7* and *ans#8* than V1, respectively. It is interesting to note that no lycopene contents were detected when the expression of BnaA03.ANS was interferenced (**Figure 7**).

The full-length cDNA fragment of *BnaA03.ANS* was also overexpressed under the control of petal-specific promoter *XY355* (Fu et al., 2018), but no obvious color change appeared with *BnaA03.ANS* was overexpressed by 144-fold or 440-fold (**Figure 6C**). Carotenoids and anthocyanin profiles showed that cyanidin (3.15-fold), neoxanthin (1.35-fold), and lycopene (1.18-fold) were high accumulated, while, pelargonidin (-85.7%), petunidin (-83.8%), lutein (-27.6%), and β-carotene (-47.7%) were lower in *ANS-OE* than in V3. However, the petal colors did not obviously differ from yellow (**Figure 7**).

DISCUSSION

Colorful flowers are one of the most important traits of many ornamental flowering plants and have gained more and more attention from breeders (Nishihara and Nakatsuka, 2011). Flower colors in rapeseed range from white, orange to yellow, and other rare colors such as pink, purple, and red also have been reported recently, which greatly enhanced its ornamental value (Fu et al., 2018). However, the underlying mechanisms of how genes and metabolites regulate color formation in red petals have not been elucidated thoroughly (Yin et al., 2019). In this study, the DAMs and DEGs in unopened small petals (stage 1) and unopened large petals (stage 4) of red color rapeseed (V1) compared with those in white (V2) and yellow (V3) were identified, aiming at understanding the molecular mechanism on the regulation of red petal color formation in rapeseed.

Metabolomic Analysis Revealed That Flavonoids May Contribute to Red Color Formation

The result of the principal component analysis of metabolomic research showed that two groups of red petals in rapeseed (V1-1 and V1-4) exhibited obvious separation from white and yellow ones (**Figures 2A,B**), which reflected that accumulation levels of metabolites in red petals have a huge difference in compared with white and yellow petals. Further analysis revealed that a total of 46+2+103 DAMs in positive mode and 20+73 DAMs in negative mode (**Figure 2**) were co-differentially accumulated in Co-DAMs-T1 (co-DAMs in V1 vs. V2 and V1 vs. V3 at stage 1), Co-DAMs-T4 (co-DAMs in V1 vs. V2 and V1 vs. V3 at stage 4), and group V1-1 vs. V1-4 (co-DAMs in V1 at stage 1 vs. V1 in stage 4), and these DAMs were considered as key metabolites that influence red petal formation and the differentiation from white and yellow petals. Among them, most DAMs were belonging to flavonoids (**Figures 3A,B**). Flavonoids, as the most important pigment, generate the widest spectrum of colors. Chen et al. (2012) reported that the flavonoid composition of white flower chrysanthemum only contained flavonols and flavones, while anthocyanins were detected in pink flowers. He et al. (2011) found that anthocyanins were presented in red and orange samples of *lycoris longituba* while no anthocyanins were detected within white and yellow. In this study, the heatmap of DAMs accumulation levels revealed that the accumulation level in red petals (V1-1 and V1-4) was higher than in white and yellow ones, with flavonoids being the most categories, which indicated the importance of flavonoids in red pigment formation compared with white and yellow.

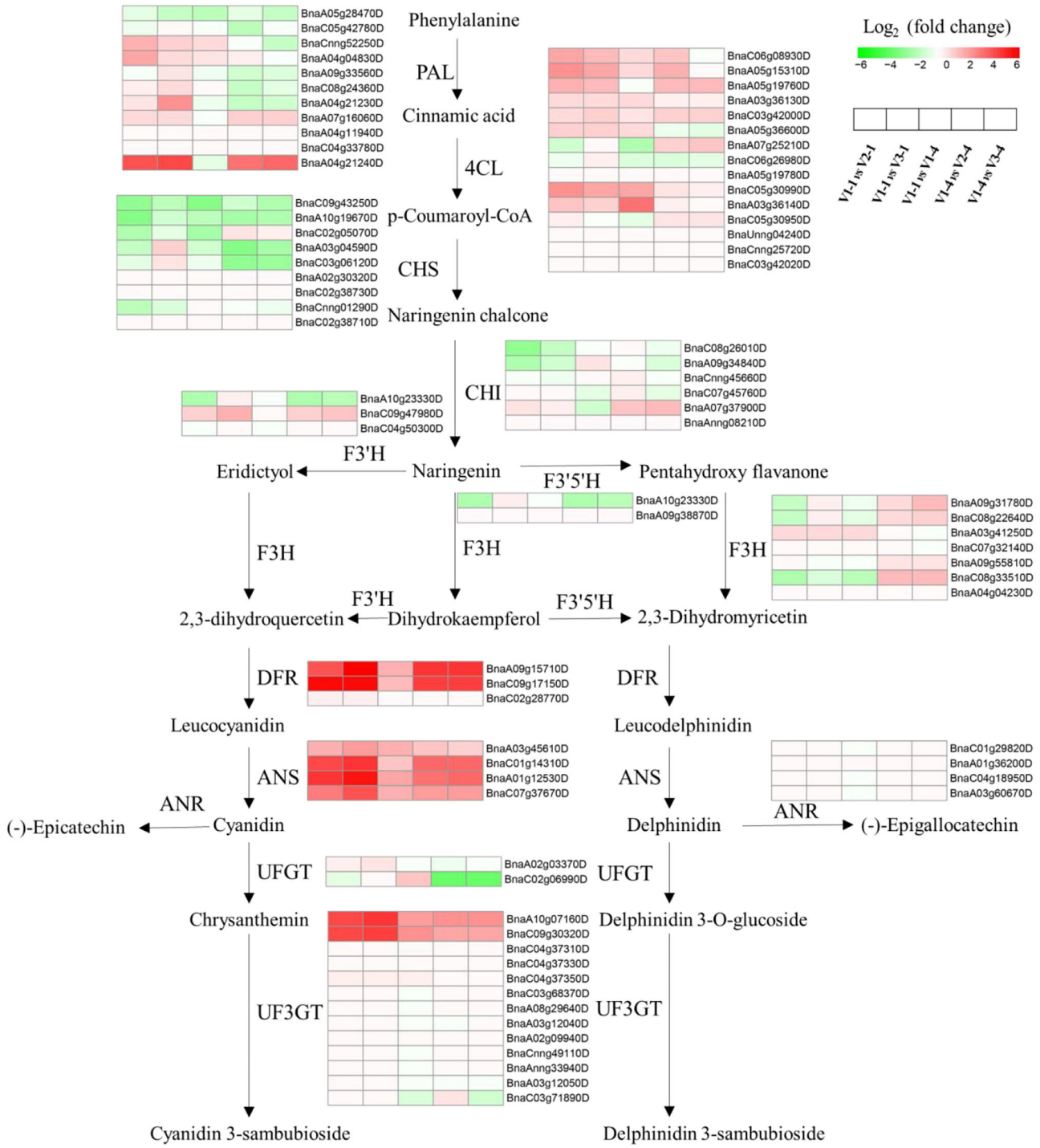


FIGURE 5 | Overview of the transcriptional regulation of anthocyanin biosynthetic genes. Heatmaps represent the $\log_2$-fold change value of V1-1 vs. V2-1, V1-1 vs. V3-1, V1-1 vs. V1-4, V1-3 vs. V2-4, and V1-4 vs. V3-4, respectively. PAL, phenylalanine ammonia lyase; 4CL, 4-coumarate co ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3'H, flavonoid 3'-hydroxylase; F3H, flavanone 3-hydroxylase; F3'5'H, flavonoid-3'5'-hydroxylase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; ANR, anthocyanidin reductase; UFGT, uridine diphosphate glucose-flavonoid glucosyltransferase; UF3GT, UDP: flavonoid 3-O-glucosyltransferase.

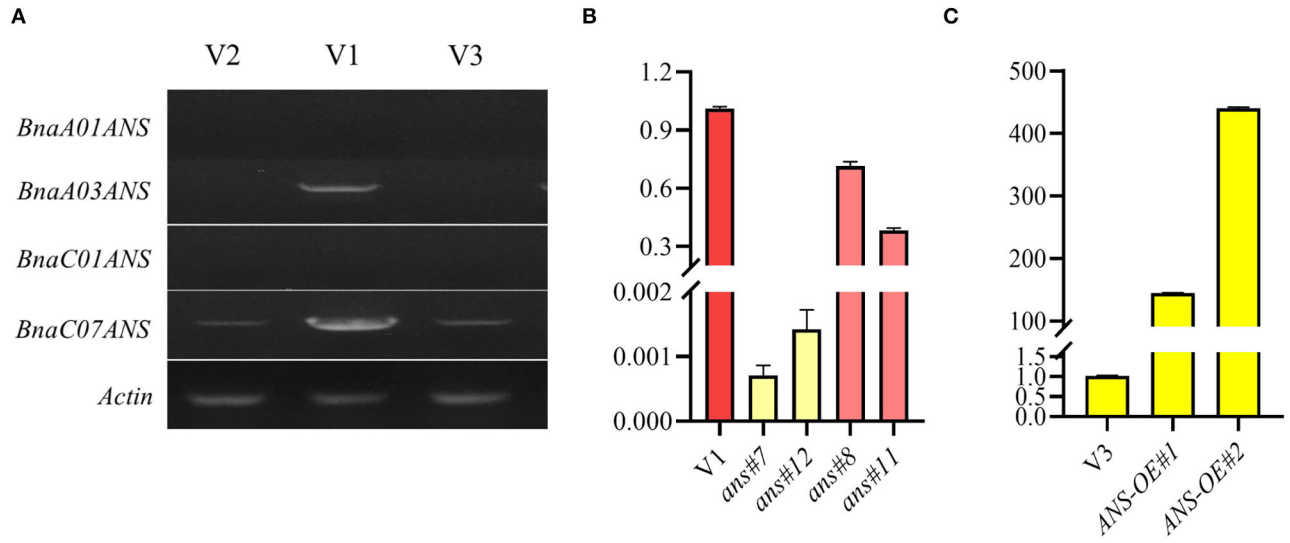


FIGURE 6 | Identification of the expression level of 4 ANS copies, and the *BnaA03.ANS* expression validation in transgenic lines. **(A)** Identification of the expression level of 4 ANS copies by semi-RT-PCR; **(B)** qRT-PCR validation of *BnaA03.ANS* expression level in RNAi lines compared with the wild type (V1). **(C)** qRT-PCR validation of *BnaA03.ANS* expression level in overexpression lines compared with the wild type (V3).

Identification of RNA-seq Data Showed Anthocyanin Biosynthesis Pathway and MYB, bHLH Transcription Factors Were Largely Differentiated in V1, V2, and V3

Venn diagrams of RNA-seq revealed that a total of 1,588 (163+127+1,298) genes (**Figure 4**) were co-differentially expressed in group Co-DEGs-T1 (co-DEGs in V1 vs. V2 and V1 vs. V3 at stage 1), group Co-DEGs-T4 (co-DEGs in V1 vs. V2 and V1 vs. V3 at stage 4), and group V1-1 vs. V1-4 (co-DEGs in V1 at stage 1 vs. V1 at stage 4), and these co-DEGs were considered as key genes that influence the red petal color formation and the differentiation from white and yellow petal colors. Anthocyanins are formed by various sugars and anthocyanidins, and the diversity of colors is highly related to anthocyanin composition and content according to a previous study (Li et al., 2003). Genes involved in the anthocyanin biosynthesis pathway that function on petal colors have been well-reported, such as *CHS* in *Malus crabapple* (Sun et al., 2015), *CHI* in asters and cyclamen (Nishihara et al., 2005), *F3H* in carnation (Owens et al., 2008), *DFR* in *Saussurea* (Li et al., 2012), and *ANS* in lisianthus flowers (Shimizu et al., 2011), which demonstrated the importance of anthocyanin biosynthesis pathway on petal color formation. As a report, red color formation, as the result of anthocyanin accumulation, is controlled through the coordination of genes that encode the enzymes involved in the anthocyanin pathway (Lai et al., 2020). For example, in radish of Cruciferae, genes for the anthocyanin biosynthesis, namely, *PAL*, *C4H*, *4CL*, *CHS*, *CHI*, *F3H*, *DFR*, and *ANS*, were identified by traditional manners and second-generation sequencing (Muleke et al., 2017; Sun et al., 2018). Among them, *RsF3H*, *RsF3'H1*, *RsCHS3*, *RsANS*, and particularly *RsUFGT*, were highly correlated with the anthocyanin contents in the flesh of red radish (Muleke et al., 2017). However, when both the skin and flesh were factored into account, only the expression of *RsDFR* and *RsANS* were correlated with the anthocyanin contents (Park et al., 2011). These results indicated that late structural genes involved in the flavonoid pathway are specifically involved in anthocyanin biosynthesis. According to our research, flavonoids were also highly accumulated from metabolomics data, and the genes and pathway involved in anthocyanin biosynthesis were drawn against these Co-DEGs-red (163+127+1,298). According to **Figure 5**, genes showed a diverse difference between red with white and yellow, while *ANS*, *DFR*, and *UF3GT* were greatly highly expressed in red petals than in white and yellow, and these results were further validated by qRT-PCR (**Supplementary Figure 2**).

In addition to these structural genes, transcription factors also play a role in petal color formation. MYB, bHLH, and WD40 families are reported to be the three major types of transcription factors that are involved in petal coloration and also regulate anthocyanin synthesis directly or indirectly (Zhang et al., 2003; Ramsay and Glover, 2005), for instance, the bleaching of gentian due to mutation of *GtMYB3* (Nakatsuka et al., 2008), and also the positive regulation function of *GhMYB10* and *LhMYB6* on anthocyanin accumulation in gerbera and lily (Roosa et al., 2008; Yamagishi et al., 2010). Overexpression of an R2R3-MYB transcription factor, *RsMYB1*, resulted in the increasing production of red flowers in radish, and by ectopic overexpression of the *OvPAP2* (*Orychophragmus violaceus*) gene, Fu et al. (2018) successfully produced red anthers and petals in yellow rapeseed (Lim et al., 2016; Fu et al., 2018). As for bHLH, in *Arabidopsis*, three bHLH transcription factors, GL3, TT8, and EGL3 were reported to participate in the biosynthesis of flavonoids (Nesi et al., 2000; Baudry et al., 2006). A recent report showed that the coexpression of RSMYB1 and RsTT8

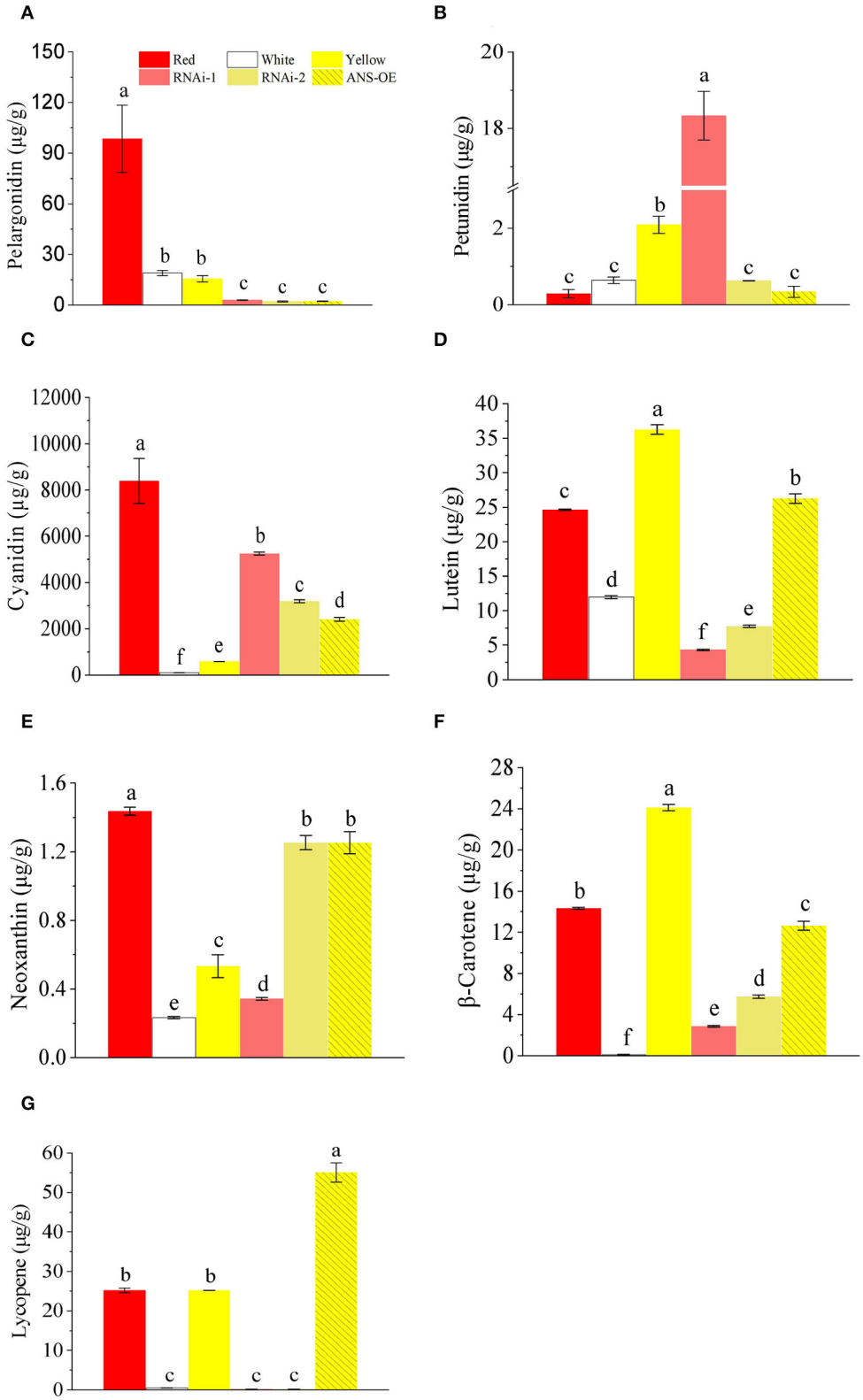


FIGURE 7 | Comparison of anthocyanins and carotenoids contents in petals of red (V1), white (V2), yellow (V3), RNAi lines, and overexpression lines. **(A)** pelargonidin; **(B)** petunidin; **(C)** cyanidin; **(D)** lutein; **(E)** neoxanthin; **(F)** β-carotene; **(G)** lycopene. Letters above the columns represent a significant difference, $p \leq 0.05$. Error bars represent SD values.

in tobacco leaves remarkably increased the accumulation of anthocyanins, indicating the partnership of RsTT8 and RsMYB1 in anthocyanin biosynthesis (Lim et al., 2017). To unveil the mechanisms of red color formation, the two-most important regulators involved in anthocyanin biosynthesis were isolated, and the expression module of MYB and bHLH were drawn in **Supplementary Figure 2**, which showed huge differences in red petals than in white and yellow at stage 1 or stage 2, reflecting the importance of MYB and bHLH on red color formation in rapeseed.

RNA Interference of *BnaA03.ANS* Transferred Petal Color From Raspberry Red to Beige Red or Zinc Yellow, While no Obvious Petal Color Was Changed When *BnaA03.ANS* Was Overexpressed

According to metabolomics and RNA-seq data, aiming at flavonoids, we selected the 4 copies of *ANS* genes that played a key role in the anthocyanin biosynthesis pathway, and semi-RT-PCR was performed. The results showed that band of *BnaA03.ANS* in red petals exhibited an obvious difference from yellow and white petals compared with the other 3 copies, which indicates the importance of *BnaA03.ANS*. Disruption of *BnaA03.ANS* in red rapeseed led to converting the petal color from raspberry red to beige red when 28.5% or 71.7% of the expression level of *BnaA03.ANS* was suppressed, while the color changed to zinc yellow, when 99.93% or 99.86% of the expression level was suppressed. The result of anthocyanins and carotenoids contents analysis revealed that pelargonidin, cyanidin, lutein, neoxanthin, β-carotene, and lycopene contents were significantly inhibited (**Figure 7**), which indicated the importance of *BnaA03.ANS* on red petal formation. However, with the overexpression of *BnaA03.ANS* under XY355 promoter in yellow rapeseed, no change of petal color was detected, though the pigments showed a different change with cyanidin, neoxanthin, and lycopene content highly accumulated while pelargonidin, petunidin, lutein, and β-carotene were decreased in overexpression lines than control (yellow rapeseed). According to a previous report, even flowers of the same color exhibited distinct pigment profiles depending on the species and variety, for example, in yellow tomato flowers, neoxanthin and violaxanthin comprise the two major carotenoids, whereas 9-cis-violaxanthin is the predominant component in yellow rose flowers (Ariizumi et al., 2014; Wan et al., 2019; Liu et al., 2020). Therefore, further researches still need to be investigated.

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 at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA848086.

AUTHOR CONTRIBUTIONS

SH, HL, and LJ designed the experiment. PH, HL, BL, YR, and LH performed the experiment. PH and SH wrote the manuscript. All authors read and agreed to the final manuscript.

FUNDING

This work was supported by the earmarked fund for the China Agriculture Research System (CARS-12), the Zhejiang Natural Science Foundation (Y21C13014), the National Natural Science Foundation (32130076), the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02064), and the Zhejiang Key Laboratory of Digital Dry Land Crops (2022E10012).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.940765/full#supplementary-material

Supplementary Figure 1 | KEGG analysis of co-DEGs relative to red pigment formation.

Supplementary Figure 2 | Heatmap of bHLH and MYB transcription factors among co-DEGs relative to red pigment formation.

Supplementary Figure 3 | qRT-PCR identification of genes relative to anthocyanin biosynthesis.

Supplementary Table 1 | Primers used in this research.

Supplementary Table 2 | Basic results of RNA-seq data.

REFERENCES

Ariizumi, T., Kishimoto, S., Kakami, R., Maoka, T., Hirakawa, H., Suzuki, Y., et al. (2014). Identification of the carotenoid modifying gene *PALE YELLOW PETAL 1* as an essential factor in xanthophyll esterification and yellow flower pigmentation in tomato (*Solanum lycopersicum*). *Plant J.* 79, 453–465. doi: 10.1111/tpj.12570

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, 768–779. doi: 10.1111/j.1365-313X.2006.02733.x

Cao, H., Zhang, J., Xu, J., Ye, J., Yun, Z., Xu, Q., et al. (2012). Comprehending crystalline b-carotene accumulation by comparing engineered cell models and the natural carotenoid-rich system of citrus. *J. Exp. Bot.* 63, 4403–4417. doi: 10.1093/jxb/ers115

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I., 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

Chen, S. M., Li, C. H., Zhu, X. R., Deng, Y. M., Sun, W., and Wang, L. S. (2012). The identification of flavonoids and the expression of genes of anthocyanin biosynthesis in the chrysanthemum flowers. *Biol. Plant.* 56, 458–464. doi: 10.1007/s10535-012-0069-3

Cheynier, V., Comte, G., Davies, K. M., Lattanzio, V., and Martens, S. (2013). Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Bioch.* 72, 1–20. doi: 10.1016/j.plaphy.2013.05.009

Fu, D. H., Jiang, L. Y., Masons, A. S., Xiao, M. L., Zhu, L. R., Li, L. Z., et al. (2016). Research progress and strategies for multifunctional rapeseed: a case

study of China. *J. Integ. Agric.* 15, 1673–1684. doi: 10.1016/S2095-3119(16)61384-9

Fu, W. Q., Chen, D. Z., Pan, Q., Li, F. F., Zhao, Z. G., Ge, X. H., et al. (2018). Production of red-flowered oilseed rape via the ectopic expression of *Orychophragmus violaceus OvPAP2*. *Plant Biotech. J.* 16, 367–380. doi: 10.1111/pbi.12777

He, Q., Shen, Y., Wang, M., Huang, M., Yang, R., and Zhu, S. (2011). Natural variation in petal color in Lycoris longituba revealed by anthocyanin components. *PLoS ONE* 6, 22098. doi: 10.1371/journal.pone.0022098

Heller, W., Forkmann, G., Britsch, L., and Grisebach, H. (1985). Enzymatic reduction of (+)-dihydroflavonols to flavan-3 4-cis-diols with flower extracts from *Matthiola incana* and its role in anthocyanin biosynthesis. *Planta* 165, 284–287. doi: 10.1007/BF00395052

Kazuma, K., Noda, N., and Suzuki, M. (2003). Flavonoid composition related to petal color in different lines of *Clitoria ternatea*. *Phytochemistry* 64, 1133–1139. doi: 10.1016/S0031-9422(03)00504-1

Lai, B., Cheng, Y. Y., Liu, H., Wang, Q. X., Wang, Q., Wang, C. L., et al. (2020). Differential anthocyanin accumulation in radish taproot: importance of *RsMYB1* gene structure. *Plant Cell Rep.* 39, 217–226. doi: 10.1007/s00299-019-02485-z

Li, H. H., Qiu, J., Chen, F. D., Lv, X. F., Fu, C. X., Zhao, D., et al. (2012). Molecular characterization and expression analysis of dihydroflavonol 4-reductase (DFR) gene in Saussurea medusa. *Mol. Biol. Rep*. 39, 2991–2999. doi: 10.1007/s11033-011-1061-2

Li, M. R., Chen, J. T., Sun, Z. J., Chen, Y. Z., and Li, H. Q. (2003). Advances in molecular breeding of ornamental plants. *J. Trop. Subtrop. Bot.* 11, 87–92. doi: 10.3969/j.issn.1005-3395.2003.01.017

Lim, S., Kim, D., Kim, J. K., Lee, J., and Ha, S. (2017). A radish basic helix-loop-helix transcription factor, RsTT8 acts a positive regulator for anthocyanin biosynthesis. *Front. Plant Sci.* 8, 1917. doi: 10.3389/fpls.2017.01917

Lim, S., Song, J., Kim, D., Kim, J. K., Lee, J., Kim, Y., et al. (2016). Activation of anthocyanin biosynthesis by expression of the radish R2R3-MYB transcription factor gene *RsMYB1*. *Plant Cell Rep.* 35, 641–653. doi: 10.1007/s00299-015-1909-3

Liu, Y. J., Ye, S. H., Yuan, G. G., Ma, X. W., Heng, S. P., Yi, B., et al. (2020). Gene silencing of *BnaA09.ZEP* and *BnaC09.ZEP* confers orange color in *Brassica napus* flowers. *Plant J.* 104, 932–949. doi: 10.1111/tpj.14970

Long, C. C., Flint, J. A., and Lepper, P. A. (2011). Insect attraction to wind turbines: does colour play a role? *Eur. J. Wildlife Res.* 57, 323–331. doi: 10.1007/s10344-010-0432-7

Morris, W., Ducreux, L., Griffiths, D., Stewart, D., Davies, H., and Taylor, M. (2004). Carotenogenesis during tuber development and storage in potato. *J. Exp. Bot.* 55, 975–982. doi: 10.1093/jxb/erh121

Muleke, E. M., Fan, L., Wang, Y., Xu, L., Zhu, X., Zhang, W., et al. (2017). Coordinated regulation of anthocyanin biosynthesis genes confers varied phenotypic and spatial-temporal anthocyanin accumulation in radish (*Raphanus sativus* L.). *Front Plant Sci.* 8, 1243. doi: 10.3389/fpls.2017.01243

Nakatsuka, A., Izumi, Y., and Yamagishi, M. (2003). Spatial and temporal expression of chalcone synthase and dihydroflavonol 4-reductase genes in the Asiatic hybrid lily. *Plant Sci.* 166, 759–767. doi: 10.1016/S0168-9452(03)00254-1

Nakatsuka, T., Haruta, K. S., Pitaksutheepong, C., Abe, Y., Kakizaki, Y., and Yamamoto, K. (2008). Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. *Plant Cell Physiol.* 49, 1818–1829. doi: 10.1093/pcp/pcn163

Nakatsuka, T., Nishihara, M., Mishiba, K., and Yamamura, S. (2005). Temporal expression of flavonoid biosynthesis related genes regulates flower pigmentation in gentian plants. *Plant Sci.* 168, 1309–1318. doi: 10.1016/j.plantsci.2005.01.009

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* 12, 1863–1878. doi: 10.1105/tpc.12.10.1863

Nikolov, L. A. (2019). Brassicaceae flowers: diversity amid uniformity. *J. Exp. Bot.* 70, 2623–2635. doi: 10.1093/jxb/erz079

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

Nishihara, M., Nakatsuka, T., and Yamamura, S. (2005). Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. *FEBS Lett.* 579, 6074–6078. doi: 10.1016/j.febslet.2005.09.073

Owens, D. K., Crosby, K. C., Runac, J., Howard, B. A., and Winkel, B. S. (2008). Biochemical and genetic characterization of *Arabidopsis* flavanone 3β-hydroxylase. *Plant Physiol. Biochem.* 46, 833–843. doi: 10.1016/j.plaphy.2008.06.004

Park, N. I., Xu, H., Li, X., Jang, I. H., Park, S., Ahn, G. H., et al. (2011). Anthocyanin accumulation and expression of anthocyanin biosynthetic genes in radish (*Raphanus sativus*). *J. Agric. Food Chem.* 59, 6034–6039. doi: 10.1021/jf200824c

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011

Roosa, A. E., Laitinen, M., Ainasoja, M., Teeri, T. H., and Elomaa, P. (2008). Identification of target genes for a MYB-type anthocyanin regulator in *Gerbera hybrida*. *J. Exp. Bot.* 59, 3691–3703. doi: 10.1093/jxb/ern216

Rosati, C., Cadic, A., Duron, M., Ingouff, M., and Simoneaub, P. (1999). Molecular characterization of the anthocyanidin synthase gene in Forsythia x intermedia reveals organ-specific expression during flower development. *Plant Sci.* 149, 73–79. doi: 10.1016/S0168-9452(99)00146-6

Sagawa, J. M., Stanley, L. E., LaFountain, A. M., Frank, H. A., Liu, C., and Yuan, Y. W. (2016). An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers. *New Phytol.* 209, 1049–1057. doi: 10.1111/nph.13647

Shimizu, K., Ohnishi, N., Morikawa, N., Ishigami, A., Otake, S., Rabah, I. O., et al. (2011). A 94-bp deletion of anthocyanidin synthase gene in acyanic flower lines of lisianthus [*Eustoma grandiflorum* (Raf.) Shinn.]. *J. Japan Soc. Hortic. Sci.* 80, 434–442. doi: 10.2503/jjshs1.80.434

Sun, L. L., Gao, W., Zhang, M. M., Li, C., Wang, A. G., Su, Y. L., et al. (2014). Composition and antioxidant activity of the anthocyanins of the fruit of *Berberis heteropoda* Schrenk. *Molecules* 19, 19078–19096. doi: 10.3390/molecules191119078

Sun, W., Li, C., Wang, L., Dai, S., and Xu, Y. (2009). Anthocyanins present in flowers of *Senecio cruentus* with different colors. *Acta Hortic. Sin.* 36, 1775–1782.

Sun, W., Meng, X., Liang, L., Jiang, W., Huang, Y., He, J., et al. (2015). Molecular and biochemical analysis of chalcone synthase from Freesia hybrid in flavonoid biosynthetic pathway. *PLoS ONE* 10, e0119054. doi: 10.1371/journal.pone.0119054

Sun, Y., Wang, J., Qiu, Y., Liu, T., Song, J., and Li, X. (2018). Identification of 'Xinlimei' radish candidate genes associated with anthocyanin biosynthesis based on a transcriptome analysis. *Gene* 657, 81–91. doi: 10.1016/j.gene.2018.03.001

Tai, D., Tian, J., Zhang, J., Song, T., and Yao, Y. (2014). A Malus crabapple chalcone synthase gene, *McCHS*, regulates red petal color and flavonoid biosynthesis. *PLoS ONE* 9, e110570. doi: 10.1371/journal.pone.0110570

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x

Veitch, N. C., and Grayer, R. J. (2008). Flavonoids and their glycosides including anthocyanins. *Nat. Prod. Rep.* 25, 555–611. doi: 10.1039/b718040n

Vignolini, S., Moyroud, E., Hingant, T., Banks, H., Rudall, P. J., Steiner, U., et al. (2015). The flower of *Hibiscus trionum* is both visibly and measurably iridescent. *New Phytol.* 205, 97–101. doi: 10.1111/nph.12958

Wan, H., Yu, C., Han, Y., Guo, X., Luo, L., Pan, H., et al. (2019). Determination of flavonoids and carotenoids and their contributions to various colors of rose cultivars (*Rosa* spp.). *Front. Plant Sci.* 10, 123. doi: 10.3389/fpls.2019.00123

Wellmann, F., Griesser, M., Schwab, W., Martens, S., Eisenreich, W., Matern, U., et al. (2006). Anthocyanidin synthase from *Gerbera hybrida* catalyzes the conversion of (+)-catechin to cyanidin and a novel procyanidin. *FEBS Lett.* 580, 1642–1648. doi: 10.1016/j.febslet.2006.02.004

Xu, P., Cao, S., Hu, K., Wang, X., Huang, W., Wang, G., et al. (2017). Trilocular phenotype in *Brassica juncea* L. resulted from interruption of *CLAVATA1* gene homologue (*BjMc1*) transcription. *Sci. Rep.* 7, 3498. doi: 10.1038/s41598-017-03755-0

Yamagishi, M., Shimoyamada, Y., Nakatsuka, T., and Masuda, K. (2010). Two R2R3-MYB genes homologs of petunia AN2 regulate anthocyanin biosynthesis

in flower tepals tepal spots and leaves of Asiatic hybrid lily. *Plant Cell Physiol.* 51, 463–474. doi: 10.1093/pcp/pcq011

Yin, N. W., Wang, S. X., Jia, L. D., Zhu, M. C., Yang, J., Zhou, B. J., et al. (2019). Identification and characterization of major constituents in different-colored rapeseed petals by UPLC-HESI-MS/MS. *J. Agric. Food Chem.* 67, 11053–11065 doi: 10.1021/acs.jafc.9b05046

Zhang, B., Liu, C., Wang, Y. Q., Yao, X., Wang, F., Wu, J. S., et al. (2015). Disruption of a *CAROTENOID CLEAVAGE DIOXYGENASE 4* gene converts flower colour from white to yellow in *Brassica* species. *New Phytol.* 206, 1513–1526. doi: 10.1111/nph.13335

Zhang, F., Gonzalez, A., Zhao, M. Z., Payne, C. T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* 130, 4859–4869. doi: 10.1242/dev.00681

Zhang, J., Wang, L., Gao, J., Li, S., Xu, Y., Li, C., et al. (2011). Identification of anthocyanins involving in petal coloration in *Chaenomeles speciosa* cultivars. *Acta Hortic. Sin.* 38, 527–534.

Zhao, D. Q., and Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Front. Plant Sci.* 6, 261. doi: 10.3389/fpls.2015.00261

Zhao, D. Q., Tao, J., Han, C. X., and Ge, J. T. (2012). Flower color diversity revealed by differential expression of flavonoid biosynthetic genes and flavonoid accumulation in herbaceous peony (*Paeonia lactiflora* Pall.). *Mol. Biol. Rep.* 39, 11263–11275. doi: 10.1007/s11033-012-2036-7

Zhou, Y., Wang, H., Gilmer, S., Whitwill, S., Keller, W., and Fowke, L. C. (2012). Control of petal and pollen development by the plant cyclin-dependent kinase inhibitor ICK1 in transgenic *Brassica* plants. *Planta* 215, 248–257. doi: 10.1007/s00425-002-0752-2

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: 04 April 2022
doi: 10.3389/fpls.2022.855486

Transcriptome and Small RNA Sequencing Reveal the Mechanisms Regulating Harvest Index in *Brassica napus*

Chao Zhang[1,2,3†], Wei Chang[1,2†], Xiaodong Li[1,2], Bo Yang[1,2], Liyuan Zhang[1,2], Zhongchun Xiao[1,2], Jiana Li[1,2,4] and Kun Lu[1,2,4*]*

[1] Chongqing Rapeseed Engineering Research Center, College of Agronomy and Biotechnology, Southwest University, Chongqing, China, [2] Academy of Agricultural Sciences, Southwest University, Chongqing, China, [3] Oil Research Institute of Guizhou Province, Guizhou Academy of Agricultural Sciences, Guiyang, China, [4] Engineering Research Center of South Upland Agriculture, Ministry of Education, Chongqing, China

OPEN ACCESS

Edited by:
Chunyu Zhang,
Huazhong Agricultural University, China

Reviewed by:
Libin Zhang,
Huazhong University of Science and Technology, China
Jinjin Jiang,
Yangzhou University, China
Maoteng Li,
Huazhong University of Science and Technology, China

***Correspondence:**
Jiana Li
ljn1950@swu.edu.cn
Kun Lu
drlukun@swu.edu.cn
orcid.org/0000-0003-1370-8633

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

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

Received: 15 January 2022
Accepted: 22 February 2022
Published: 04 April 2022

Citation:
Zhang C, Chang W, Li X, Yang B, Zhang L, Xiao Z, Li J and and Lu K (2022) Transcriptome and Small RNA Sequencing Reveal the Mechanisms Regulating Harvest Index in Brassica napus. Front. Plant Sci. 13:855486. doi: 10.3389/fpls.2022.855486

Harvest index (HI), the ratio of harvested seed weight to total aboveground biomass weight, is an economically critical value reflecting the convergence of complex agronomic traits. HI values in rapeseed (*Brassica napus*) remain much lower than in other major crops, and the underlying regulatory network is largely unknown. In this study, we performed mRNA and small RNA sequencing to reveal the mechanisms shaping HI in *B. napus* during the seed-filling stage. A total of 8,410 differentially expressed genes (DEGs) between high-HI and low-HI accessions in four tissues (silique pericarp, seed, leaves, and stem) were identified. Combining with co-expression network, 72 gene modules were identified, and a key gene *BnaSTY46* was found to participate in retarded establishment of photosynthetic capacity to influence HI. Further research found that the genes involved in circadian rhythms and response to stimulus may play important roles in HI and that their transcript levels were modulated by differentially expressed microRNAs (DEMs), and we identified 903 microRNAs (miRNAs), including 46 known miRNAs and 857 novel miRNAs. Furthermore, transporter activity-related genes were critical to enhancing HI in good cultivation environments. Of 903 miRNAs, we found that the bna-miR396–Bna.A06SRp34a/Bna.A01EMB3119 pair may control the seed development and the accumulation of storage compounds, thus contributing to higher HI. Our findings uncovered the underlying complex regulatory network behind HI and offer potential approaches to rapeseed improvement.

Keywords: harvest index, transcriptome, miRNA, regulatory network, *Brassica napus*

INTRODUCTION

Harvest index (HI), formerly known as coefficient of economics, refers to the ratio of the economic yield to the biological yield of the crop at harvest (Donald, 1962), which can also be understood as the ratio of seeds harvested from a plant to all the above-ground biomass produced. HI thus reflects the distribution of crop assimilates products in economic yield organs and vegetative organs, and also indicates the patency of crop photosynthate transport from "source" organs

to "sink" organs (Bennett et al., 2011, 2012), and HI has clear agronomic implications. Since the first Green Revolution, the increase in the yield of major crops such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), and barley (*Hordeum vulgare*) has been mainly due to the increase in the HI (Austin et al., 1980). The allotetraploid crop rapeseed (*Brassica napus*) is cultivated worldwide for oil, which is extracted from its oil-rich seeds. Luo et al. (2015) examined the *B. napus* genetic architecture of HI using 35,791 high-throughput single nucleotide polymorphisms (SNPs) genotyped by the Illumina Brassica SNP60 Bead Chip in an association panel with 155 accessions, and a total of nine SNPs on the C genome were identified to be significantly associated with HI; they explained 3.42% of the phenotypic variance in HI. Based on Brassica SNP60 Bead Chip, a natural population (NP) containing 520 materials was used to perform genome-wide association study (GWAS) for traits related to HI. Combined with transcriptomic sequencing (RNA-seq) of materials with high and low HI, candidate genes involved in photosynthesis, inflorescence, and silique development were identified (Lu et al., 2016). In addition, in *B. napus*, it was also found that the *BnaA02.TB1* (*BnaA02g14010D*) that regulates lateral branch development and the *BnaA02.GW2* (*BnaA02g18890D*) that regulates grain weight may be related to HI (Chao et al., 2019); *BnaDwf.C9* (*BnaC09g20450D*) and *BnaC04.BIL1* (*BnaC04g41660D*) can affect HI by regulating plant height (Wang X. D. et al., 2020; Yang et al., 2021). These previous studies are of great worth for helping us to elucidate the genetic mechanism of HI in *B. napus*. As seed yield is part of the HI numerator, HI increases with grain yield and has, therefore, received widespread attention in breeding programs (Hay, 1995; Sadras and Lawson, 2011; Beche et al., 2014). Unfortunately, the HI of rapeseed remains much lower than that of other major crops, such as rice, wheat, maize (*Zea mays*), and soybean (*Glycine max*) (Hay, 1995; Unkovich et al., 2010). The HI for rapeseed ranges from 0.05 to 0.42 (Lu et al., 2016), which is far below the theoretical biological limit for HI of ∼0.60 in grain crops (Foulkes et al., 2009), indicating the potential to further increase HI. Therefore, improving HI in rapeseed varieties is a major objective for breeders.

Green plant tissues such as leaves and silique pericarps (SP) constitute "source" organs that are photosynthetically active and provide photoassimilates to "sink" organs like grains and seeds *via* translocation (designated here as "flow"); HI reflects flow and the balance between source tissues and sink organs (Unkovich et al., 2010). In *B. napus*, flow is thought to be the limiting factor in the accumulation of assimilates in seeds (Shen et al., 2010; Fu and Zhou, 2013; Luo et al., 2015). Unlike many other crops in which leaves provide the main source tissue throughout seed development, in *B. napus*, SP take over the role of source tissue from senescent leaves, providing nutrients to sustain the growth of the seed after silique formation (Diepenbrock, 2000; Bennett et al., 2011). Several critical sugar transporter gene families have been identified in *B. napus*, such as *SUCROSE TRANSPORTER* (*BnSUC*), *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER* (*BnSWEET*), and *MONOSACCHARIDE TRANSPORTER* (*BnMST*); transporters encoded by these genes may influence HI positively during the development of siliques by supplying more sugars to the seeds (Jian et al., 2016; Zhang et al., 2020). In addition, the *Arabidopsis thaliana AMINO ACID PERMEASE 2* (*AAP2*) gene encodes an amino acid co-transporter that transports amino acids to the embryo *via* the seed pod vascular system during the seed-filling phase (Hirner et al., 1998), suggesting that SP may supply not only sugars but also other compounds. However, the underlying molecular regulatory mechanism behind the translocation of photoassimilates and other nutrients during the seed-filling phase is poorly understood in *B. napus*.

Harvest index reflects the relative allocation of resources to vegetative and reproductive organs and can be influenced by various environmental factors, including water supply (Gajića et al., 2018), temperature stress (Prasad et al., 2006), and nitrogen (N) fertilization (Amanullah and Shah, 2010). For crop species such as maize, shortening the vegetative growth period reduces the accumulation of photosynthetic products in vegetative organs and raises HI (Hütsch and Schubert, 2017). One method employed to modulate the length of the vegetative period relies on the genetic manipulation of photoperiod sensitivity (Li et al., 2020). In addition, enhancing crop adaptability to external stimuli can also result in significant increases in harvestable products and thus contribute to a higher HI. For example, the rice basic leucine zipper transcription factor bZIP58 induces the expression of seed storage protein genes and starch biosynthetic genes, but its transcripts also undergo alternative splicing in response to heat stress. However, more heat-tolerant rice varieties showed limited alternative splicing and higher expression of seed storage protein genes under heat stress, thus contributing to heat tolerance during grain filling (Xu et al., 2020). In rice, a dominant mutation in the *DENSE AND ERECT PANICLE 1* (*DEP1*) gene displayed insensitivity to N during vegetative growth, which increased N uptake and assimilation and thus improved HI and grain yield at moderate levels of N fertilization (Sun et al., 2014). However, which genes regulate HI by promoting reproductive growth and mediating responses to external stimuli remain largely unknown in *B. napus*.

MicroRNAs (miRNAs) are 20–24-nt single-stranded non-coding RNA molecules that act as key post-transcriptional regulators of gene expression (Voinnet, 2009; Budak and Akpinar, 2015). Extensive studies have revealed that miRNAs play diverse roles during plant development and adjust complex traits in crops (D'Ario et al., 2017; Tang and Chu, 2017). For example, Arabidopsis miR159 lowers the transcript levels of the *MYB DOMAIN PROTEIN 33* (*MYB33*) gene, whose encoded transcription factor normally negatively regulates miR156, thereby modulating vegetative phase change (Guo et al., 2017). Similarly, miR164 induces cleavage of *NAM*, *ATAF1/2*, and *CUC2* (*NAC2*) gene transcripts, which act as negative regulators of drought tolerance in rice (Fang et al., 2014). In Arabidopsis, miR397b regulates *LACCASE 4* (*LAC4*) and influences silique numbers and silique length (Wang et al., 2014). Moreover, rice miR396d forms a molecular bridge between *BRASSINAZOLE-RESISTANT 1* (*BZR1*) and *GROWTH REGULATING FACTORs* (*GRFs*); *OsBZR1* induces the expression of miR396d, which, in turn, represses *OsGRF* transcript accumulation to modulate plant architecture and grain yield (Tang et al., 2018).

The advent of next-generation sequencing technologies has paved the way to cost-effective and highly efficient methods to identify miRNA–mRNA regulatory networks related to complex traits in crops. For instance, embryo development in peanut (*Arachis hypogaea*) is highly sensitive to calcium deficiency in the soil, but the cause for the resulting embryo abortion was unknown. However, an integrated miRNA and mRNA profiling (RNA-seq) analysis revealed that a number of miRNA-mediated regulatory networks affecting seed/embryo development, cell division, cell proliferation, and plant hormone signaling all participated in peanut embryo abortion under calcium deficiency (Chen Z. Y. et al., 2019). Joint RNA-Seq and miRNA profiling analyses further established that miRNAs involved in nitrogen-related pathways regulated the thickness of a pod canopy in *B. napus* (Chen Z. Y. et al., 2019). To date, limited information is available pertaining to miRNA-mediated regulatory networks related to HI.

To better understand the regulatory networks underlying the establishment of HI in *B. napus* during the seed-filling phase, we determined the mRNA and miRNA transcriptome landscape in SP and seeds during the seed-filling stage in plants grown at two locations, Chongqing and Yunnan. We identified potential gene clusters involved in the regulation of HI, with predicted roles in transporter activity and responses to environmental signals. Furthermore, we discovered several miRNA-mediated regulatory networks in SP and seeds. These results contribute to uncovering the complex regulatory networks behind HI and offering potential solutions for its improvement *via* genetic engineering or crop breeding.

MATERIALS AND METHODS

Plant Materials and Field Trials

The *B. napus* accessions YC24 (SWU47), YC52 (Zhongshuang11), and YC46 (Ningyou12) with stability HI for two consecutive years (2012–13 and 2013–14) were selected from 520 accessions in previous study (Lu et al., 2016) and grown in a randomized block design with three replications at Chongqing Beibei (CQ, 29°45′ N, 106°22′ E, 238.6-m altitude) and Yunnan Lincang (YN, 23°43′ N, 100°02′ E, 1819.5-m altitude) during the 2015–2016 growing season. For simplicity, we denoted the planting location first (CQ or YN), followed by the accession number (24, 52, or 46) to distinguish experimental groups. Planting conditions were as previously described (Lu et al., 2016). Plant materials used for RNA-seq and small RNA-seq of growing status are shown in **Supplementary Figure 1**. We collected HI-related phenotypic traits, such as HI, biomass yield per plant (BY), seed yield per plant (SY), stem dry weight (ST), canopy biomass yield (CBY), according to the calculation method used previously (Lu et al., 2017), and the HI, BY, and SY of CQ24 and CQ46 at 2016 had been shown in our previous study (Zhang et al., 2020).

RNA Isolation and Library Preparation

For each accession grown at CQ and YN, we harvested seeds (ZS) and SP from the main inflorescence at 30 days after flowering (denoted as 30ZS and 30SP, respectively); there are also leaves (Le, main leaves at 30 days after flowering) and stems (St, main steam at 30 days after flowering) of the same period. For each sample, we collected two biological replicates, each harvested from five independent plants.

Total RNA of each sample was extracted using a CTAB method (Lu et al., 2008). RNA degradation and DNA contamination were monitored by gel electrophoresis on 1% agarose gels. RNA purity was confirmed on a NanoPhotometer spectrophotometer (IMPLEN, CA, United States), and RNA concentration was measured using the Qubit RNA Assay Kit with a Qubit 2.0 Fluorometer (Life Technologies, CA, United States). RNA integrity was assessed using the RNA 6000 Nano Assay Kit on an Agilent Bioanalyzer 2100 system (Technologies, CA, United States). After quality-control, we sent 48 RNA samples for mRNA sequencing and selected 12 samples (including 30SP, 30ZS) for small RNA sequencing to Novogene Corporation (Beijing, China) for library construction and sequencing, as previously described (Chen H. et al., 2019).

Identification of Differentially Expressed Genes

We evaluated the quality of RNA-Seq data with the Trimmomatic software (v.0.36) (Bolger et al., 2014). We removed adapter sequences and low-quality reads from the raw data. All clean reads were then mapped to the *B. napus* reference genomev.4.1[1] using the STAR program (v.2.5.3) (Dobin et al., 2013) with default parameters. We then mapped the aligned reads to RNA features using the feature Counts function of Subread (v.1.6.0) (Liao et al., 2014). We identified [differentially expressed gene (DEGs)] with the gene counts generated above with the edgeR package (Robinson et al., 2010) to import, organize, filter, and normalize the data with a false discovery rate (FDR) of 1%. We quantified relative gene expression as fragments per kilobase of exon model per million mapped reads (FPKM) by Cuffdiff within Cufflinks (Trapnell et al., 2012). After calculating their expression fold-change, we identified genes with FDR-adjusted q-value ≤ 0.05 and absolute $\log_2$ (fold-change) ≥ 1 as DEGs.

Identification of Differentially Expressed Conserved and Novel MicroRNAs

Clean data were obtained from small RNA sequencing libraries by removing adapter sequences from all reads, as well as reads containing over 10% Ns and low-quality (Q20 < 85%) reads from the raw data. We extracted potential small RNAs 18–30 nt in length and mapped them to the *B. napus* reference genome using Bowtie2 (v.2.2.9) (Langmead et al., 2009). Mapped reads were further mapped to the released *B. napus* miRNAs in miRBase22 (v.22.1)[2] using basic local alignment sequence tool for nucleotides (BLASTN) with an *E*-value cutoff of $\leq$ 1e-5, which identified known miRNAs. We used the Rfam database (v.14.4)[3] to remove ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear

[1]http://www.genoscope.cns.fr/brassicanapus/data/

[2]http://www.mirbase.org/

[3]http://rfam.xfam.org/

RNAs (snRNAs), and small nucleolar RNAs (snoRNAs). In order to improve the accuracy and efficiency of comparison, we retained only the longest sequence from the remaining reads and merged them with the sequences in miRBase22 to become the final *B. napus* miRNA dataset. After removing the reads already classified as known miRNAs, we employed miRDeep-P (v.1.1.4) (Yang and Li, 2011) to predict potential novel miRNAs, allowing for 2 bp 3′ overhangs. We named these potential novel miRNAs based on the chromosome they map to and their starting nucleotide position on that chromosome.

To identify differentially expressed miRNAs (DEMs), we normalized miRNA expression across all samples using Salmon (v.0.11.3) (Patro et al., 2017) to obtain the expression of transcript per million (TPM) based on the normalization formula: normalized expression = (actual miRNA counts/total number of mapped reads) × 1,000,000. We calculated the associated fold-change, *p*-values, and *q*-values with in-house Perl scripts. miRNAs with FDR-adjusted *q*-value ≤ 0.05 and absolute $\log_2$ (fold-change) ≥ 1 were determined to be DEMs.

MicroRNA Target Prediction

We employed the psRNATarget server (v. 2017)[4] to predict the target genes for all known and novel miRNAs that were differentially expressed between comparable groups, with default parameters.

Gene Function Clustering Analysis

Functional annotation was performed by using BLASTX to compare *B. napus* and Arabidopsis proteins. We used the online tool agriGO (gene ontology analysis toolkit and database for agricultural community, v.2.0)[5] (Tian et al., 2017) for GO enrichment analysis. We determined GO classifications by submitting the gene sequences to the BLAST4ID tool of agriGO to obtain the corresponding *B. napus* locus ID, and then running GO term enrichment analysis.

Weighted Gene Co-expression Network Analysis

Weighted gene co-expression network analysis (WGCNA) package in R was designed in 2008 for helping users create weighted correlation network modules and identify key genes associated with traits in interesting modules (Langfelder and Horvath, 2008). FPKM values of genes were $\log_2$ (FPKM + 1) transformed for further calculation of correlation coefficient, determination of gene modules, construction of co-expression network, and correlation analysis of modules and phenotypic traits. In the process of analysis, the soft thresholding power was determined using the pickSoftTreshold function based on the scale-free topology model fit (R^2) > 0.9; the automatic blockwiseModules network construction approach was applied to obtain the highly correlated modules, with the following parameters: power = 7; maxBlockSize = 30,000; TOM-type = unsigned; miniModuleSize = 30; reassignThreshold = 0, mergeCutHeight = 0.25. The co-expression networks were displayed using Cytoscape (v.3.5.1) (Maere et al., 2005).

Validation of Transcript Levels by Quantitative Reverse Transcription-PCR

The same RNA samples prepared for mRNA and miRNA sequencing libraries were used for cDNA synthesis and quantitative reverse transcription-PCR (qRT-PCR) detection. qRT-PCR was performed as described previously (Qu et al., 2015). We used gene-specific primers obtained from qPrimerDB[6] (Lu et al., 2018). Relative transcript levels were normalized to the *B. napus* housekeeping genes *UBIQUITIN-CONJUGATING ENZYME 21* (*Bna.UBC21*) and *Bna.ACTIN7* (Qu et al., 2016).

For the validation of miRNAs, we added a poly (A) tail and performed reverse transcription from 2-μg RNA in 20-μl reaction volume using the miRcute miRNA First-Strand cDNA Synthesis Kit (Tiangen, Beijing, China). We then diluted the cDNAs 10-fold for RT-qPCR analysis, using the miRcute miRNA qPCR Detection Kit (SYBR) (Tiangen, Beijing, China). We carried out PCR reaction solutions containing 1-μl (∼10 ng) diluted cDNA, 10-μl 2- × -miRcute miRNA premix, a 0.4-μl forward primer (10 μM), a 0.4-μl reverse primer (10 μM), and 8.2-μl ddH_2O on a BIO-RAD CFX96 Real-Time System (BIORAD, United States). Cycle conditions were: 95°C for 15 min, followed by five cycles of 94°C for 20 s, 65°C for 30 s, and 72°C for 34 s, and then 40 cycles of 94°C for 20 s and 60°C for 30 s. All reactions were performed in triplicate, with the *B. napus* U6 snRNA as internal control. Relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). All above-mentioned qRT-PCR assays were performed according to MIQE guidelines (Bustin et al., 2010). Three independent biological replicates, each with three technical replicates, were implemented for each sample.

RESULTS

Physiological Characteristics of Accessions With Different Harvest Index

In the context of this study, we planted the *B. napus* accessions YC24, YC52, and YC46 at the Beibei, Chongqing (CQ), and Lincang, Yunnan (YN) locations during the 2015–2016 growing season. All phenotypic values are provided in **Table 1**; in order to distinguish experimental groups, they were named as planting location first (CQ or YN) plus the accession number (24, 52, or 46). We collected HI-related phenotypic traits, such as HI, biomass yield per plant (BY), seed yield per plant (SY), stem dry weight (ST), canopy biomass yield (CBY). We noticed that the HI for all accessions significantly increased at the YN location when compared to that measured at CQ, especially accessions YC24 and YC52. In addition, YC24 displayed a significantly higher HI than the YC52 and YC46 accessions at both locations (**Table 1**); therefore, YC24 was regarded as a high HI accession, and YC52 and YC46 as low HI accessions. We used these three accessions to elucidate the molecular mechanism of HI.

[4]http://plantgrn.noble.org/psRNATarget/

[5]http://systemsbiology.cau.edu.cn/agriGOv2/

[6]https://biodb.swu.edu.cn/qprimerdb/

TABLE 1 | Phenotypic values of the three varieties measured in Chongqing and Yunnan.

	CBY (g)	SY (g)	ST (g)	BY (g)	HI (%)
CQ24	68.8 ± 4.0cd	29.6 ± 3.8cd	50.8 ± 6.1bc	119.6 ± 9.5bc	24.8 ± 2.7c
CQ52	75.9 ± 3.4c	25.5 ± 0.3d	54.2 ± 6.6bc	130.1 ± 4.0b	19.6 ± 0.5d
CQ46	56.0 ± 5.4d	15.0 ± 1.6e	42.6 ± 3.0c	98.6 ± 8.4c	15.2 ± 0.9e
YN24	132.2 ± 12.7a	62.8 ± 6.2a	61.1 ± 14.8b	193.3 ± 27.5a	32.6 ± 1.4a
YN52	120.6 ± 3.5a	52.8 ± 2.4b	66.2 ± 6.4b	186.8 ± 5.5a	28.2 ± 0.8b
YN46	102.8 ± 17.1b	35.3 ± 7.3c	86.8 ± 10.1a	189.7 ± 26.4a	18.5 ± 1.7d

The data are means ± SD; columns with different letters indicate significant differences based on Duncan's multiple range tests at $p < 0.05$.

mRNA Sequencing Data Analysis Uncovers Harvest Index-Related Differentially Expressed Genes

We collected 48 samples for RNA extraction and subsequent deep sequencing of the transcriptome (RNA-Seq) analysis: ZS, SP, Le, and St in biological duplicates from all three accessions at the two geographic locations. After removal of low-quality reads from the raw data, we retained 40.29∼52.68 million clean reads; 83.97 to 95.23% of which were mapped to the *B. napus* reference genome (except SP in CQ24 Sample 1, which shows just 61.67%). Sample correlation analysis used FPKM to emphasize the high degree of correlation between biological replicates (**Supplementary Table 1**). The raw sequencing data were deposited in the BIG Data Center under BioProject accession No. PRJNA602979.

We identified DEGs across SP samples; we detected 2,303 DEGs (1,212 upregulated and 1,091 downregulated) in CQ24 vs. CQ52; 4,680 DEGs (2,472 upregulated and 2,208 downregulated) in CQ24 vs. CQ46; 1,103 DEGs (641 upregulated, 462 downregulated) in YN24 vs. YN52; 1,904 DEGs (1,092 upregulated, 812 downregulated) in YN24 vs. YN46 (**Supplementary Table 2**). To distinguish HI-related DEGs between high- and low-HI lines, we generated a four-way Venn diagram showing the extent of overlap between the four comparisons listed above. This analysis highlighted 756 DEGs (393 upregulated and 363 downregulated) in SP between high- and low-HI accessions grown at the CQ location, and 343 DEGs (203 upregulated and 140 downregulated) between high- and low-HI accessions grown at the YN location. A subset of 146 DEGs (83 upregulated and 63 downregulated) was identified in SP samples collected at both locations and across all accessions (**Figures 1A,B**).

We repeated the same analysis with the RNA-Seq data obtained from the ZS samples; we identified 2,192 DEGs (1,481 upregulated, 711 downregulated) in CQ24 vs. CQ52; 2,776 DEGs (1,430 upregulated and 1,346 downregulated) in CQ24 vs. CQ46; 850 DEGs (562 upregulated, 288 downregulated) in YN24 vs. YN52; 1,963 DEGs (1,281 upregulated, 682 downregulated) in YN24 vs. YN46 (**Supplementary Table 2**). The corresponding four-way Venn diagram showed that 438 DEGs (236 upregulated and 202 downregulated) were shared by high- and low-HI accessions grown at the CQ location, while 205 DEGs (149 upregulated and 56 downregulated) were common to high- and low-HI accessions grown at the YN location. Furthermore, 118 genes (83 upregulated and 35 downregulated) were differentially expressed in seed samples collected at both locations and across all accessions (**Figures 1C,D**).

Leaves are the same "source" organs as the SP; through RNA-seq data, we identified 2,500 DEGs (1,227 upregulated, 1,273 downregulated) in CQ24 vs. CQ52; 6,047 DEGs (3,019 upregulated and 3,026 downregulated) in CQ24 vs. CQ46; 4,660 DEGs (2,185 upregulated, 2,475 downregulated) in YN24 vs. YN52; 6,250 DEGs (2,833 upregulated, 3,417 downregulated) in YN24 vs. YN46 (**Supplementary Table 2**). Our 4-way Venn diagram showed that 783 DEGs (371 upregulated and 412 downregulated) were shared by high- and low-HI accessions grown at the CQ location, while 1,369 DEGs (588 upregulated and 781 downregulated) were common to high- and low-HI accessions grown at the YN location. And 552 (242 upregulated and 310 downregulated) genes were differentially expressed in Le samples collected at both locations and across all accessions (**Figures 1E,F**).

Stem as the "flow" organ, which is the limitation for the accumulation of assimilates in *B. napus* seeds (Shen et al., 2010; Fu and Zhou, 2013; Luo et al., 2015), we also analysis stem RNA-seq data; there are 1,415 DEGs (722 upregulated, 693 downregulated) in CQ24 vs. CQ52; 5,313 DEGs (2,652 upregulated and 2,761 downregulated) in CQ24 vs. CQ46; 1,143 DEGs (623 upregulated, 520 downregulated) in YN24 vs. YN52; 4,866 DEGs (2,559 upregulated, 2,307 downregulated) in YN24 vs. YN46 (**Supplementary Table 2**). And 418 DEGs (214 upregulated, 204 downregulated) in St between high- and low-HI accessions grown at the CQ location; 311 DEGs (199 upregulated, 112 downregulated) in St between high- and low-HI accessions grown at the YN location. About 227 (132 upregulated, 95 downregulated) genes were differentially expressed in St samples collected from the two locations (**Figures 1G,H**).

Overall, the transcriptome of the accessions YC24 and YC46 differed by more DEGs than when YC24 was compared to the YC52 accession, both in SP, Le, ZS, and St. In addition, SP and Le samples as "source" organs were characterized by more DEGs than St and ZS samples between high- and low-HI accessions at both locations, indicating that the regulation of HI in "source" organs might be more complicated. Finally, samples harvested at the CQ location exhibited more DEGs than at the YN location in SP, ZS, and St, except Le, suggesting that the seed-filling process might be more complex at the CQ location.

In addition, in order to compare the gene expression differences of the same material in different regions, we compared the DEGs of the same material grown in YN and CQ. The

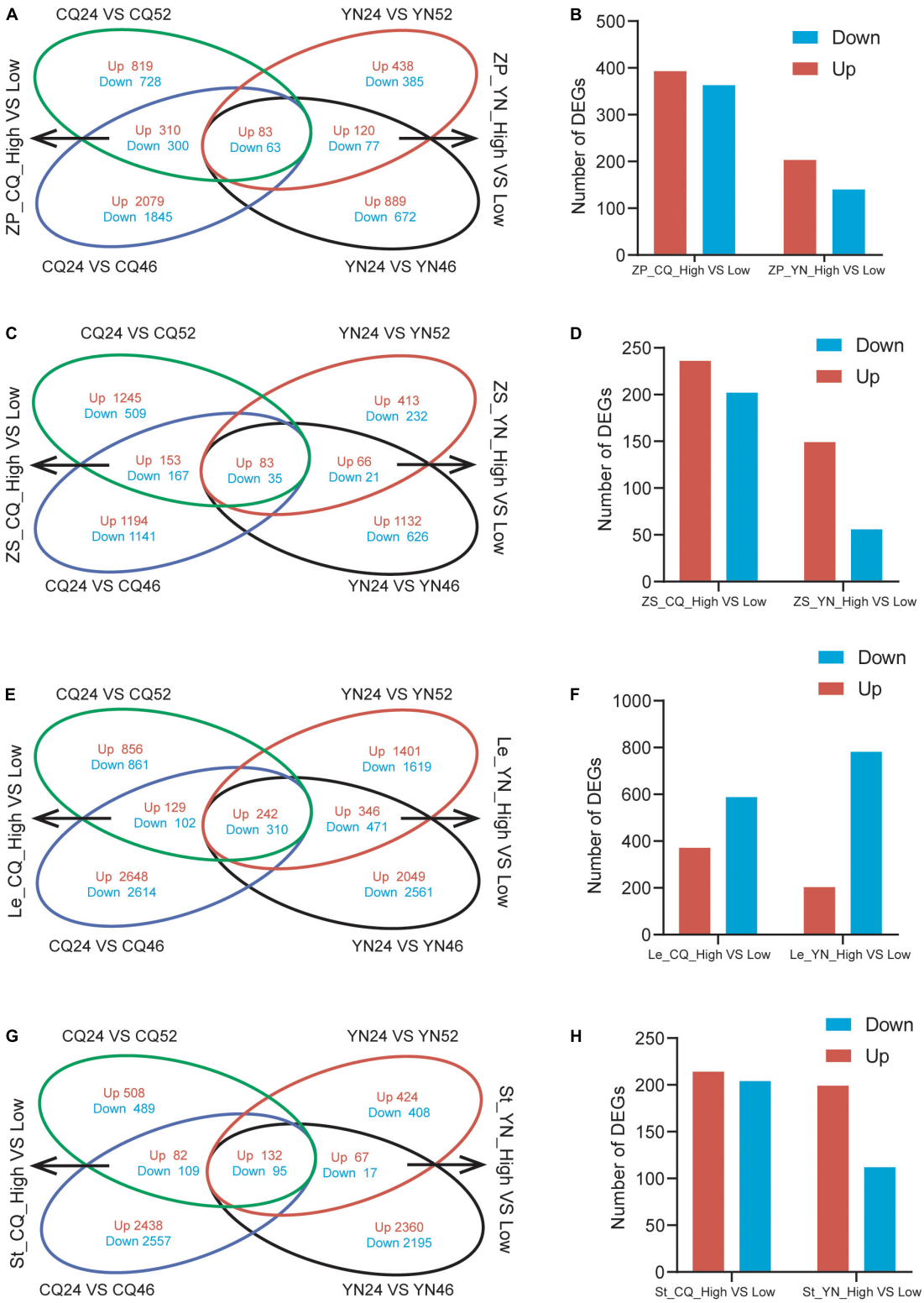


FIGURE 1 | Venn diagrams and the number of DEGs of the DEGs from high- and low-HI accessions. **(A)** A Venn diagram of DEGs between high- and low-HI accessions in ZP at CQ and YN. **(B)** The number of DEGs between high- and low-HI accessions in ZP. **(C)** A Venn diagram of DEGs between high- and low-HI accessions in Le at CQ and YN. **(D)** The number of DEGs between high- and low- HI accessions in Le. **(E)** A Venn diagram of DEGs between high- and low-HI accessions in ZS at CQ and YN. **(F)** The number of DEGs between high- and low-HI accessions in ZS. **(G)** A Venn diagram of DEGs between high- and low-HI accessions in ST at CQ and YN. **(H)** The number of DEGs between high- and low-HI accessions in ZP in ST.

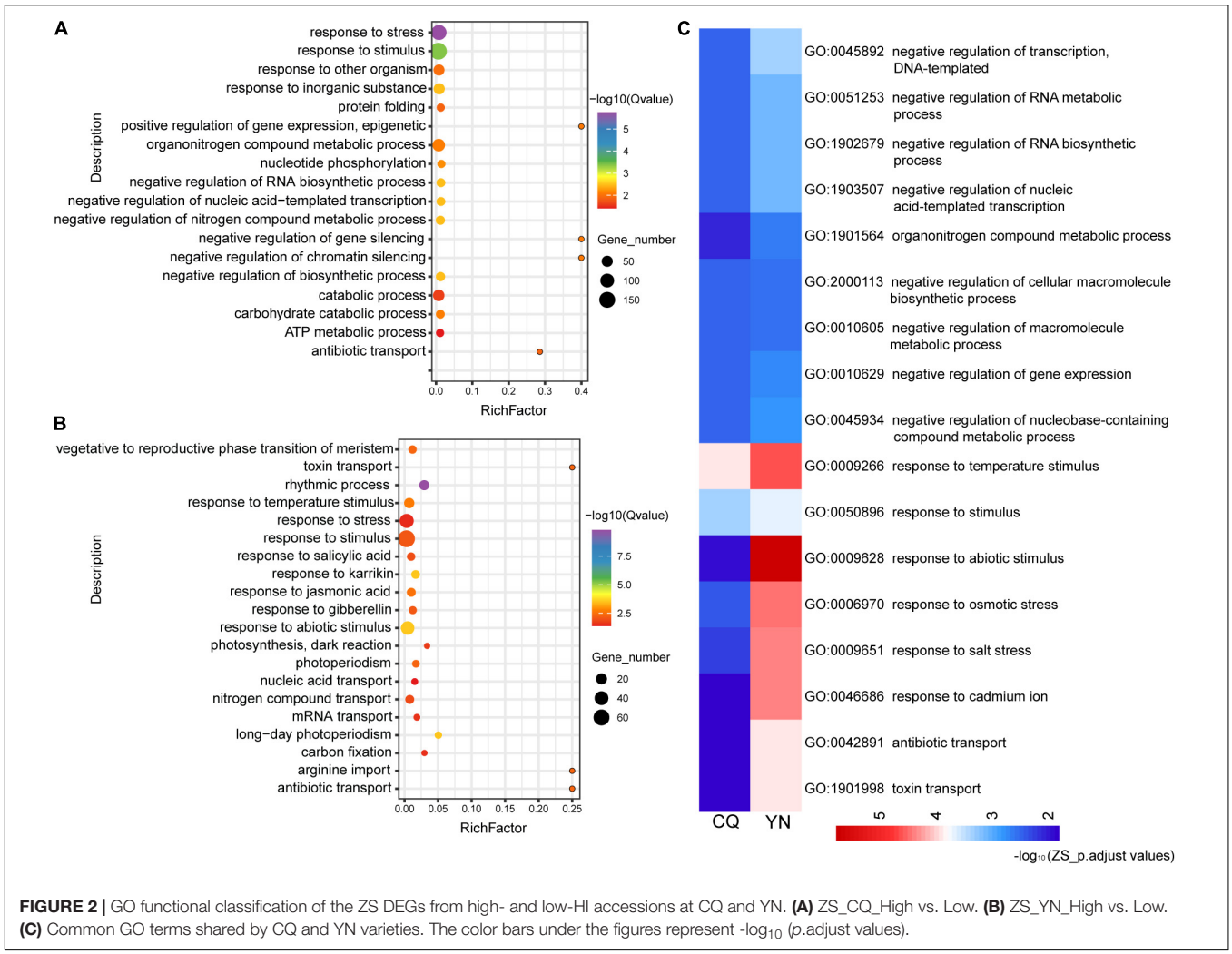


FIGURE 2 | GO functional classification of the ZS DEGs from high- and low-HI accessions at CQ and YN. **(A)** ZS_CQ_High vs. Low. **(B)** ZS_YN_High vs. Low. **(C)** Common GO terms shared by CQ and YN varieties. The color bars under the figures represent -$\log_{10}$ (*p*.adjust values).

results indicate that these DEGs among the three materials in four different plant tissues are quite different in different environments and the expression levels also varied greatly (**Supplementary Table 2**), especially in the Le of material YC24; there were 4,599 upregulated DEGs and 4,659 downregulated DEGs (**Supplementary Figure 2**). At the same time, the DEGs in the same tissue of the three materials also have differences; only 16, 36, and 34 same DEGs were found in the St, Le, and SP, and the absence of the same DEGs was found in the ZS (**Supplementary Figure 2**). Moreover, combining with **Table 1**, we found that HI in YN was significantly higher than that in CQ, indicating that HI was easily affected by environmental conditions; this is worthy of further study.

Functional Annotation and Classification of Differentially Expressed Genes

To understand the functions encoded by the DEGs identified between high- and low-HI accessions, we performed a GO enrichment. Studies showed long ago that the photosynthates of the SP are the main sources of seed yield, contributing about 2/3 of the total dry matter of the seed yield, whereas the beak of the seeds is about 8% (Leng et al., 1992), which can seriously affect HI, so we focused on SP and ZS first.

Differentially expressed genes in SP, harvested at the CQ location between high- and low-HI accessions (SP_CQ_High vs. Low), were associated with 93 significantly enriched GO terms (q-value $\leq$ 0.05), while DEGs for the equivalent samples collected at the YN location (SP_YN_High vs. Low) showed a significant enrichment for 341 GO terms; in addition, we identified 54 significantly enriched GO terms (q-value $\leq$ 0.05) for SP among DEGs shared by both CQ and YN locations (**Supplementary Figures 3A,B**). We observed a significant enrichment in several cellular component GO terms related to "chloroplasts" (GO: 0005737, GO: 0009507, GO: 0009536, GO: 0044434, GO: 0044435), and we noticed that GO terms of SP in addition to response to biological stress and a large number of abiotic stresses (GO:0009628, response to abiotic stimulus; GO:0009408, response to heat; GO:0009607, response to biotic stimulus; etc.); there are some interesting pathways enriched in "circadian rhythm" (GO:0007623), "histone H3-K36 demethylation" (GO:0070544), and "response to Karrikin" (GO:0080167), which suggest that

circadian rhythm, histone modification affected the plant HI; also, as Karrikin, a new signaling molecule participates in regulating HI (**Supplementary Figure 3C**).

In ZS samples, a similar analysis revealed 107 significantly enriched GO terms among DEGs between high- and low-HI accessions harvested at the CQ location (ZS_CQ_High vs. Low), and 59 significantly enriched GO terms among DEGs between high- and low-HI accessions harvested at the YN location (ZS_YN_High vs. Low) (**Figures 2A,B**). We identified 18 common enriched GO terms among DEGs for ZS. Biological processes-related GO terms, such as the "negative regulation of RNA metabolic process" (GO: 1902679), "negative regulation of nucleic acid-templated transcription" (GO: 1903507), "negative regulation of the RNA metabolic process" (GO: 0051253), and the "organonitrogen compound metabolic process" (GO: 1901564), were predominantly enriched in ZS samples harvested at both CQ and YN locations (**Figure 2C**). In the YN location, we noticed "circadian rhythm" (GO: 0007623) and "alternative mRNA splicing *via* spliceosome" (GO: 0000380) were identified but not identified in the CQ location. We suspected that, in ZS samples, circadian rhythm and alternative mRNA splicing may participate in the HI regulation but were also influenced by geographical differences and environmental factors.

GO enrichment analysis of St and Le DEGs in the CQ location obviously enriched in "circadian rhythm" (GO: 0007623) as the results in SP and ZS; we speculated that circadian rhythm plays an important role in regulating HI. We also found St and Le DEGs are mainly enriched in hormone-related terms, such as the "ethylene-mediated signaling pathway" (GO: 0009873), "jasmonic acid-mediated signaling pathway" (GO: 0009867) in St at the CQ location, "response to gibberellin stimulus" (GO: 0009739) in St at the YN location. In Le samples, "response to gibberellin stimulus" (GO: 0009739) and the auxin metabolic/biosynthetic process (GO: 0009850 and GO: 0009851) also enriched, suggesting that, although the leaf no longer provides most of the photosynthetic energy during silique ripening (SP provides more), it works with the St to regulate plant growth hormonally (**Supplementary Figure 4**).

Construction of Co-expression Networks

Through WGCNA, we constructed co-expression networks with all DEGs and 13 phenotypic data (**Supplementary Table 3**). This analysis yielded 72 gene modules, each represented by different colors in the output (**Figure 3**). The smallest module is the ME light coral, which included only 55 genes, and the largest module is turquoise, which included 10,008 genes. We focused on the ME light green module that is highly correlated with the HI ($r^2 = 0.61$).

GO enrichment analysis showed the ME light green module was mainly enriched in the "ubiquitin-dependent protein catabolic process" (GO: 0006511), "entrainment of the circadian clock" (GO: 0009649), "positive regulation of meiosis" (GO: 0045836), and "production of miRNAs involved in gene silencing by miRNA" (GO: 0035196). This indicated that ubiquitin modification, circadian rhythm, and gene silencing are all strong connection with HI; the results are the same with our previous DEGs analyzed. We used the maximal clique centrality method in the cytoHubba plugin of Cytoscape (v. 3.5.1) to identify hub genes in the ME light green module of interest, and we identified *BnaA07g02330D* (*Bna.A07STY46*), which is serine/threonine kinase that phosphorylates transit peptides of chloroplast and mitochondria-targeted pre-proteins and is involved in chloroplast differentiation in Arabidopsis (Giorgia et al., 2011). We speculated that *Bna.A07STY46* may participate in retarded establishment of a photosynthetic capacity to influence HI.

Identification of Differentially Expressed MicroRNAs Based on MicroRNA Sequencing

In addition to 48 mRNA libraries, we also sequenced 12 miRNA libraries of samples SP and ZS. We obtained 163,448,662 reads and retained 159,966,665 clean reads after removing low-quality reads and adapters (**Supplementary Table 4**). We then selected clean reads with a length of 18–30 nt for further analysis. As expected, reads with a length of 21, 22, and 24 nt were more abundant out of all reads. In addition, we observed a higher fraction of 21-nt reads in silique pericarp samples when compared to that in seed samples, while 24-nt reads showed the opposite pattern (**Supplementary Figure 5**). We identified 903 miRNAs from the 12 miRNA libraries, including 46 known miRNAs and 857 novel miRNAs (**Supplementary Table 5**).

Based on the selection criteria of an absolute $\log_2$ FC > 1 and a q-value < 0.05, we detected DEMs between the same comparison groups as for RNA-Seq analysis. When we compared SP_CQ_High and Low, we identified 11 known DEMs (1 upregulated and 10 downregulated) and 86 novel DEMs (60 upregulated and 26 downregulated), while the comparison of SP_YN_High with Low revealed 15 known DEMs (0 upregulated and 15 downregulated) and 80 novel DEMs (38 upregulated and 42 downregulated). A comparison between ZS_CQ_High and Low yielded 10 known DEMs (8 upregulated and 2 downregulated) and 70 novel DEMs (38 upregulated and 32 downregulated), while we identified 11 known DEMs (11 upregulated and 0 downregulated) and 65 novel DEMs (30 upregulated and 35 downregulated) from a comparison between ZS_YN_High and Low. In addition, SP samples harvested at the CQ and YN locations shared 8 known DEMs and 21 novel DEMs, and ZS samples collected at the two locations saw an overlap consisting of 8 known DEMs and 16 novel DEMs (**Supplementary Figure 6**).

Integration of Differentially Expressed Genes and Differentially Expressed MicroRNAs

To elucidate the regulatory role of DEMs between high- and low-HI accessions, we first identified the potential target genes of each DEM before combining the expression profiles of DEMs and their target genes for further analysis. We thus obtained miRNA–mRNA interaction pairs (pairs with either negatively or positively correlated expression patterns) through a comparison of high- and low-HI accessions; 130 pairs in SP were harvested at CQ, 68 pairs in SP were collected at YN, 69 pairs in ZS from the CQ location, and 23 pairs in ZS from the YN location. Overall, almost half of the miRNA–mRNA

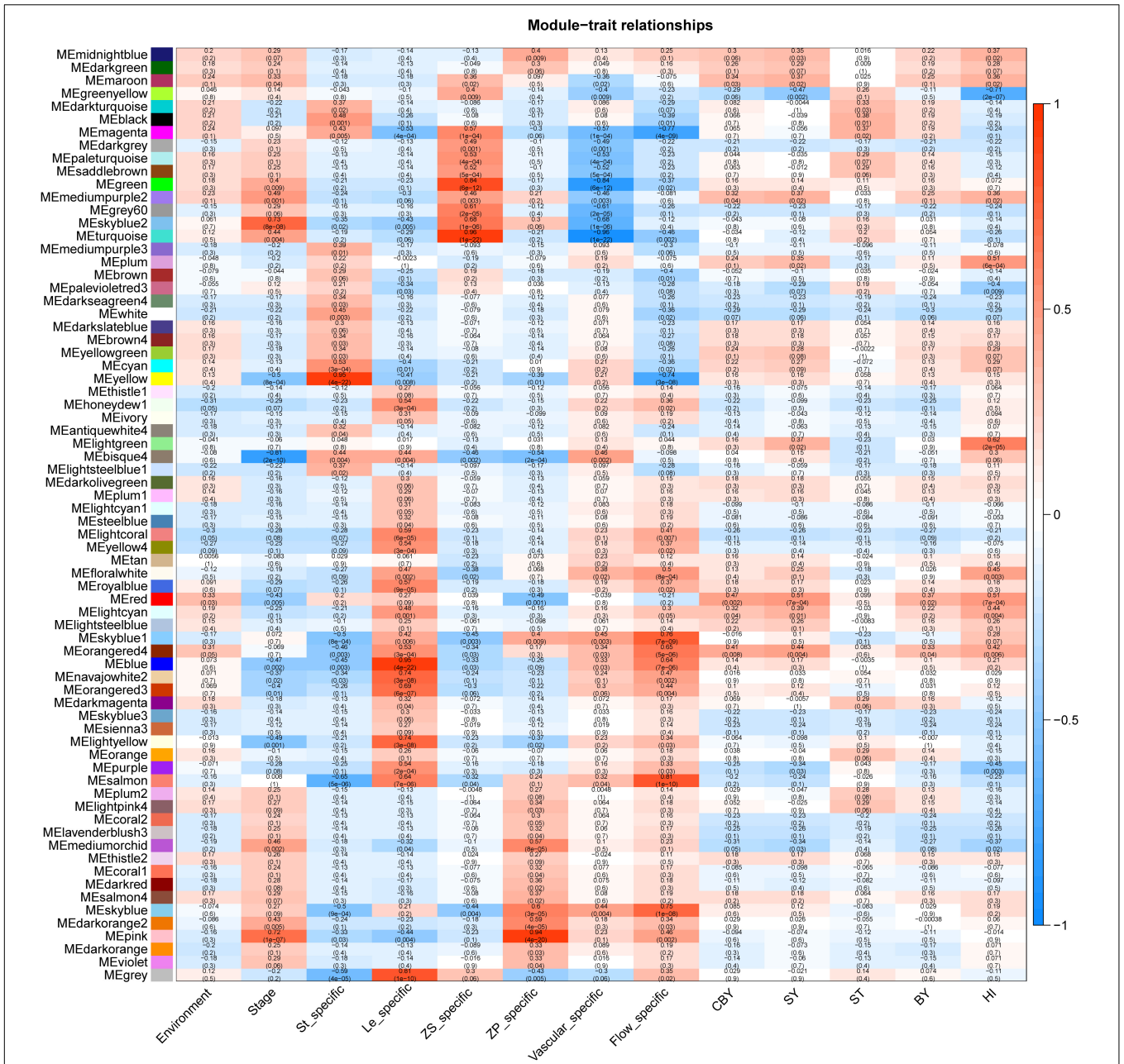


FIGURE 3 | An overview of module and traits corresponding through WGCNA.

interactions pairs showed a negative correlation, as might be expected from a true mRNA–miRNA pair involving transcript cleavage. We also noticed that several miRNAs had multiple potential target mRNAs, and multiple miRNAs that targeted a single mRNA (**Supplementary Figure 7** and **Supplementary Table 6**). For instance, the upregulated miRNA bna-miRC03_52 controlled the downregulated genes *BnaC07g39100D*, *BnaC08g32460D*, and *BnaA06g21420D*, while the upregulated miRNA bna-miRA02_2282 controlled the downregulated genes *BnaA06g21420D* and *BnaA06g39650D* in comparisons between SP_CQ_High and Low (**Supplementary Figure 7**). These observations indicate that the regulation of miRNA–mRNA pairs is complex.

We performed a GO functional annotation analysis to characterize these differentially expressed target genes. Our results showed enrichment in GO terms related to circadian rhythms, transporter activity response to stress, response to abiotic stimulus, and response to stimulus across all comparisons, with the exception of ZS_CQ_High vs. Low (**Supplementary Figure 8** and **Supplementary Table 7**). We hand-selected miRNA–mRNA pairs associated with the GO terms mentioned above, such as

the bna-miRC08_5718–*BnaA03g03740D/BnaC03g05240D* pair and the bna-miRC01_19092–*BnaC09g43920D* pair (**Figure 4A** and **Supplementary Figure 8**). We did not discover significantly enriched GO terms among miRNA–mRNA pairs from seed samples collected at the CQ and YN locations between high- and low-HI accessions. However, the bna-miR396–*BnaA06g21030D* and bna-miR396–*BnaA01g33410D* pairs were shared between the two locations (**Figure 4B** and **Supplementary Figure 8**).

Validation of mRNA and MicroRNA Expression by Quantitative Reverse Transcription-PCR

To validate the quantification of the mRNA and miRNA sequencing data presented here, we analyzed the relative transcript levels of nine randomly selected DEGs (BnaA01g26430D, *BnaA08g25340D*, *BnaA08g26300D*, *BnaC02g04730D*, *BnaC02g12960D*, *BnaC04g50590D*, *BnaC06g05910D*, *BnaC09g48250D*, and *BnaCnng27780D*) and 7 DEMs (bna-miR156a, bna-miR164a, bna-miR167a, bna-miR396, bna-miRA01_10807, bna-miRC03_27293, and bna-miRC04_29081) by qRT-PCR on the same RNA used for library construction (**Supplementary Table 8** and **Figure 5**). We observed a high degree of positive correlation between the relative transcript levels of DEGs and DEMs determined by qRT-PCR and their relative expression measured from high-throughput sequencing data. The high-throughput sequencing data used in this study were, therefore, accurate and reliable.

DISCUSSION

Harvest index is a complex agronomic trait of great economic value that depends on interactions between a plant genotype and the environment (Amanullah and Shah, 2010; Unkovich et al., 2010; Gajića et al., 2018; Chao et al., 2019). We determined that the YC24 accession showed a significantly higher HI than the YC52 accession, while the YC52 accession had a significantly higher HI than the YC46 accession at both the CQ and YN locations. These results indicated that HI differences between the YC24, YC52, and YC46 accessions were robust, validating the use of these three accessions to dissect the regulatory mechanism behind HI. YN is a high-yield crop production environment; not surprisingly, HI for the YC24, YC52, and YC46 accessions grown at this location was higher than that from the CQ location, indicating that environmental conditions can have a strong influence on HI. Thus, we carried out transcriptome sequencing to compare the gene expression profile in accessions grown in a standard HI (CQ) and high-HI (YN) environment during the seed-filling stage.

RNA-Seq and Expression Profiles of High- and Low-Harvest Index Materials

Based on gene functional annotation data, we identified significantly enriched GO terms associated with DEGs between high- and low-HI accessions. In *B. napus*, SP are photosynthetically active, and, as source tissues, they thus provide nutrients to seeds, a sink organ (Diepenbrock, 2000; Bennett et al., 2011). Furthermore, SP mediate maternal control during seed filling (Li et al., 2019). We also noticed that a large fraction of enriched GO terms were specific to the CQ location, as we did not detect them in our analysis of silique pericarp or seed samples collected at the YN location, underscoring the marked environmental sensitivity of HI regulation.

As HI represents the ratio between reproductive organs and vegetative biomass produced, regulating the timing of the phase transition from vegetative to reproductive growth will be beneficial to increasing HI (Hütsch and Schubert, 2017). Based on GO enrichment analysis, we noticed that the circadian rhythm may be associated with HI. In Arabidopsis, the core circadian rhythm comprises the proteins CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), LATE ELONGATED HYPOCOTYL (LHY), TIMING OF CAB EXPRESSION 1 (TOC1), and PSEUDO RESPONSE REGULATORS (PRRs) (Mcclung and Gutiérrez, 2010). In addition, the *CONSTANS* gene family (*CO*) controls photoperiodic flowering time together with *PRRs* (Nakamichi et al., 2007). Transcriptome analysis revealed that genes associated with circadian rhythm were potentially involved in potato (*Solanum tuberosum*) tuber formation, suggesting that the circadian rhythm may participate in photoassimilate distribution (Shan et al., 2013). In rice, *CONSTANS-like* (*OsCOL9*) not only modulated photoperiodic flowering but influenced grain number of the main panicle (Liu et al., 2016). Similarly, *TOC1* modulated chickpea (*Cicer arietinum*) seed yield per plant (Basu et al., 2019), while overexpression of Arabidopsis *PRR5* in rice delayed flowering but also significantly increased biomass (Nakamichi et al., 2020), likely due to higher expression of *OsPRR37*; the overexpression line $GCA1^{OX}$-5 showed better general combining ability of rice (Liu et al., 2015). In wheat, the *Photoperiod-1* (*Ppd-1*) mutant inactivates a *PRR* and affects paired spikelet formation (Boden et al., 2015). Moreover, our analysis highlighted additional circadian rhythm-related genes [*GLYCINE RICH PROTEIN7* (*GRP7*), *EARLY FLOWERING4* (*ELF4*), *REVEILLE 1* (*RVE1*)] as being differentially expressed between high- and low-HI lines, strongly suggesting that the circadian rhythm and rhythm-controlled gene regulation might be harnessed to modulate HI in *B. napus* in the future.

Under the high-HI environment of the YN location, we detected a number of DEGs expressed in SP that were related to transporter activity, such as *PLASMA MEMBRANE INTRINSIC PROTEIN* (*PIP*), *BILE ACID TRANSPORTER 5* (*BAT5*), *ABC2 HOMOLOG 13* (*ATH13*), *CHLORIDE CHANNEL A* (*CLC-A*), *MAJOR FACILITATOR SUPERFAMILY PROTEIN*, *NITRATE TRANSPORTER 1.7* (*NRT1.7*), *POLYOL/MONOSACCHARIDE TRANSPORTER 5* (*PMT5*), and *VACUOLAR GLUCOSE TRANSPORTER 1* (*VGT1*). PIP are aquaporins that localize to the plasma membrane and facilitate the flux of water and solutes across the plasma membrane (Kourghi et al., 2017; Wang H. et al., 2020). Multiple studies have shown that *PIP* genes can affect water balance and solute transport in plant cells (Sommer et al., 2008; Byrt et al., 2017; Macho-Rivero et al., 2018). The BAT5 transporter translocates glucosinolates, which are derived from methionine and sugars, across the chloroplast membranes

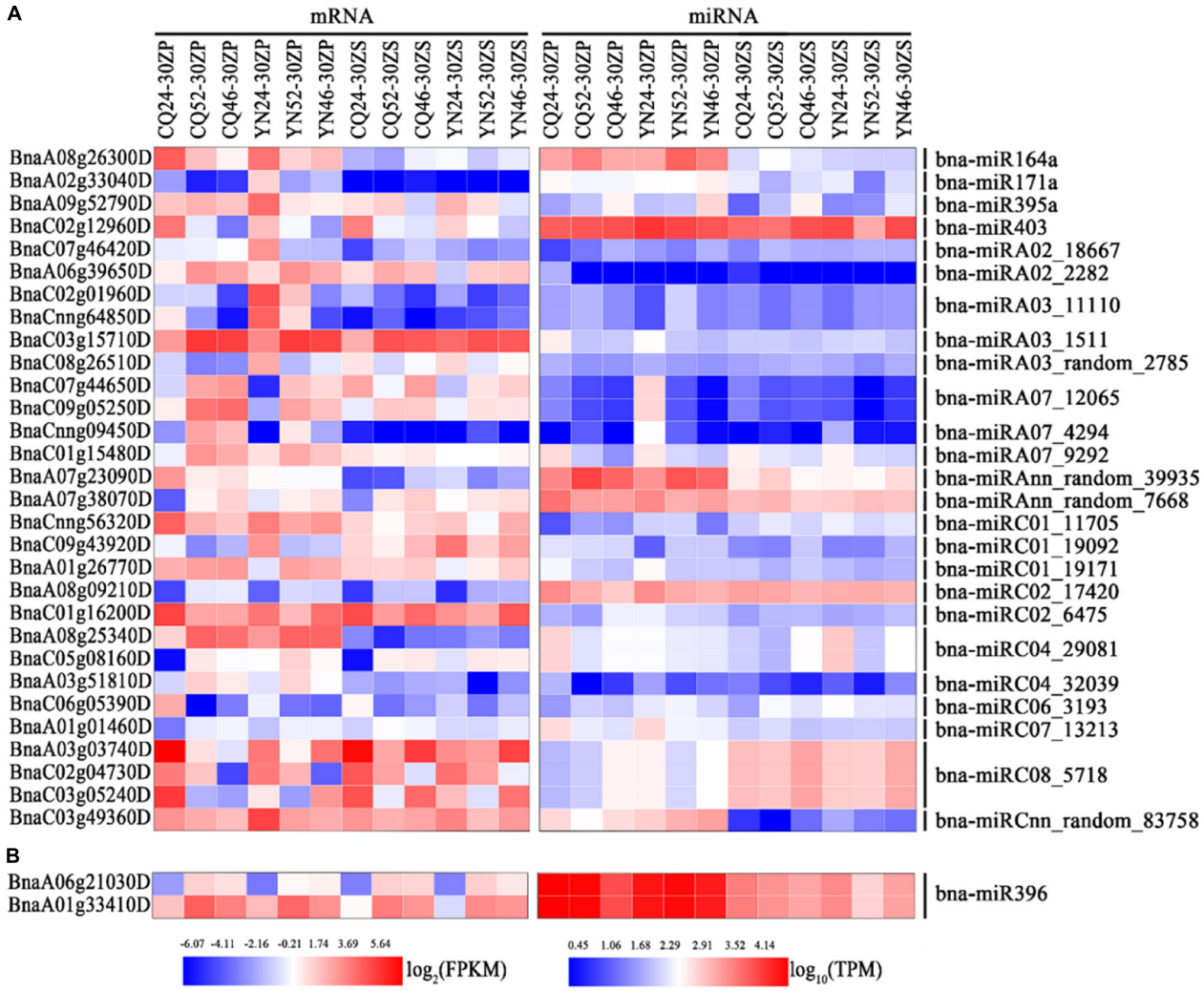


FIGURE 4 | Heatmap representation of interesting miRNA-mRNA pairs identified in silique pericarps and seeds. **(A)** Heatmap representation of miRNA-mRNA pairs-related circadian rhythm, response to stress, response to abiotic stimulus, and response to stimulus in silique pericarps. **(B)** Heatmap representation of *bna*-miR396-modulated miRNA-mRNA pairs in seeds. The color bars under the figures represent $\log_2$ (FPKM/TPM).

(Sawada et al., 2009). ATH13 affects the lipid composition of chloroplast membranes and regulates iron distribution within chloroplasts (Manara et al., 2014, 2015). CLC-A regulates the accumulation of nitrate within vacuoles and plays an important role in modifying cytosolic conditions (De Angeli et al., 2006; Manara et al., 2015; Demes et al., 2020). Additional regulators of nitrate balance and transport include members of the NITRATE TRANSPORTER1/PEPTIDE TRANSPORTER (NPF) family, such as AtNPF3.1 (At1g68570) (Yang et al., 2017) and AtNPF6.2 (At2g26690) (Tong et al., 2016); AtNRT1.7/NPF2.13 (At1g69870) also plays important roles in source-to-sink remobilization of nitrate in Arabidopsis (Fan et al., 2009; Liu et al., 2017). PMT5 can transport a wide range of linear polyols (three to six carbon backbones), cyclic polyols (myo-inositol), pyranose, furanose, hexoses, and pentoses across the plasma membrane (Klepek et al., 2005, 2010). Collectively, the upregulation of genes-encoding proteins with transporter activity in SP of high-HI accessions points to their possible involvement in facilitating water and solute fluxes (e.g., sugar, nitrate nutrients, and cations) from the mother plant to the developing seed, and thus in increasing HI. The only downregulated transporter identified in this study was *VGT1*. However, VGT1 localizes to the vacuolar membrane and mediates the transport of glucose from the cytoplasm to the vacuole (Aluri and Büttner, 2007; Büttner, 2007); a reduction in *VGT1* expression may, therefore, promote sugar translocation to sink tissues by a modulating glucose flux to the vacuole.

In seeds, GO terms, such as GO: 0045892, GO: 1902679, GO: 1903507, GO: 0051253, GO: 0045934, GO: 0010629, GO: 0010605, and GO: 1901564, were significantly enriched in samples harvested at both the CQ and YN locations. For the latter GO term (the organonitrogen compound metabolic process, GO: 1901564), the *B. napus* gene *Bna*C08g45660D homologous to *At*1g01090 (*PYRUVATE DEHYDROGENASE E1 ALPHA, PDH-E1 ALPHA*) was upregulated in high-HI

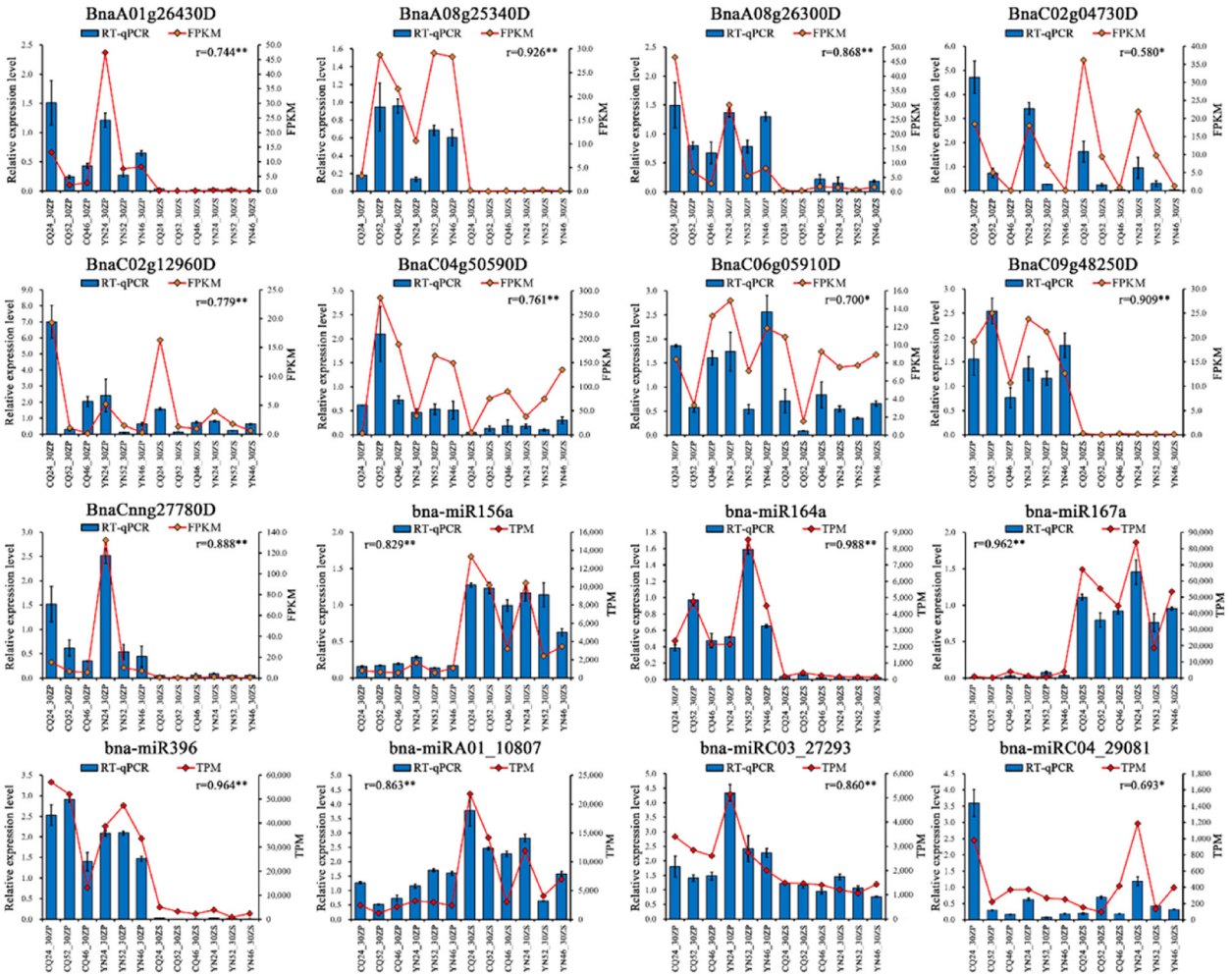


FIGURE 5 | RT-qPCR validation of DEGs and DEMs. Blue bars represent the relative expression level of RT-qPCR, and red points represent FPKM/TPM of sequencing; *r* represents Pearson correlation coefficient, *: correlation is significant at the 0.05 level, **: correlation is significant at the 0.01 level.

accessions. As *PDH-E1 ALPHA* has been shown to regulate acyl lipid metabolism (LeClere et al., 2004; Mentzen et al., 2008), we speculated that the upregulation of *Bna*C08g45660D contributes to the accumulation of storage lipids in seeds. Gene functional annotation of the GO terms GO: 0045892, GO: 1902679, GO: 1903507, GO: 0051253, GO: 0045934, GO: 0010629, and GO: 0010605 determined that the expression pattern of *Bna*A05g01050D (homologous to *At*2g46830, *CCA1*) showed the same trend at both the CQ and YN locations. Studies have shown that *CCA1* affects not only the circadian rhythm but also the accumulation of storage lipids (Kim et al., 2019). Thus, we concluded that genes that promote the accumulation of storage materials such as lipids during the seed-filling stage likely contribute to the improvement of HI.

MicroRNA-Mediated Regulatory Networks Related to Harvest Index

MicroRNAs play versatile roles in plant growth and development *via* miRNA–mRNA interaction networks. In our study, we identified abundant miRNA–mRNA interaction pairs from the comparison of silique pericarp and seed samples from high- and low-HI accessions at CQ and YN locations. Based on the functional analysis of target DEGs, we noticed enrichments for genes in silique pericarp samples harvested at both CQ and YN locations related to responses to stimulus. Thus, we selected genes related to circadian rhythm, response to stress, response to abiotic stimulus, and response to stimulus that also exhibited anti-correlations with their miRNAs for further analysis. MiR164 has been shown to control axillary meristem and floral organogenesis formation in Arabidopsis (Raman et al., 2008; Huang et al., 2012); in our study, we observed a negative correlation between bna-miR164a and *BnaA08g26300D*. *BnaA08g26300D* shows homology to the Arabidopsis gene *At1g09560* (*GERMIN-LIKE PROTEIN 5*, *GLP5*). GLP may control resource allocation between primary and lateral roots by phloem-mediated transport in Arabidopsis (Ham et al., 2012), suggesting that the regulation of *GLP5* transcript levels by bna-miR164a may adjust the balance of resources between SP and seeds though the phloem. At5g06530

(ABC TRANSPORTER GENE 22, AtABCG22) also contributes to water transpiration and drought tolerance in Arabidopsis (Kuromori et al., 2011). Hence, we propose that its *B. napus* homolog *BnaC02g01960D*, whose transcript levels are regulated by bna-miRA03_11110, might be important for the proper water balance of SP during the seed-filling phase.

The circadian rhythm gene *TOC1* contributes to energy metabolism by influencing the phase of the circadian rhythm under environmental fluctuations (Legnaioli et al., 2009; Fung-Uceda et al., 2018) and Arabidopsis; *toc1* mutants showed a modified pattern of starch mobilization under light–dark cycles (Flis et al., 2019). We hypothesized that the bna-miRA07_12065–*BnaC09g05250D* (*TOC1*) pair might similarly affect the phase of the circadian rhythm to modulate resources allocation in SP. *BnaC09g43920D* is homologous to *At5g13170* (*SENESCENCE-ASSOCIATED GENE 29*, *SAG29*, also named *SWEET15*). In this study, this gene is upregulated in high-HI SP at the YN location *via* the downregulation of bna-miRC01_19092. SWEET15, together with the transporters SWEET11 and SWEET12, mediates sucrose efflux both intracellularly and intercellularly during seed filling in Arabidopsis (Chen et al., 2015; Eom et al., 2015) and may also regulate senescence under environmental stress (Seo et al., 2011). Moreover, we noted the upregulation of several heat stress-related genes, such as *HEAT SHOCK PROTEIN* (*HSP*) *Hsp17.6CII* (*Bna*A03g03740D/*Bna*C03g05240D), *HSP26.5* (*Bna*C06g05390D), *CPHSC70-1* (*Bna*C01g16200D), and *HEAT SHOCK FACTOR 4* (*HSF4*) (*Bna*Cnng56320D), this upregulation being mediated by the downregulated miRNAs bna-miRC08_5718, bna-miRC06_3193, bna-miRC02_6475, and bna-miRC01_11705 in high-HI lines, respectively. Overall, we identified many miRNA–mRNA pairs related to environmental stress in the current study, indicating that increasing plant adaptability to the environment may adjust resources distribution and improve HI.

Similar to our observation with DEGs, numerous miRNA–mRNA pairs identified here differed between seeds and SP, indicating distinct regulatory mechanisms during the seed-filling stage. Among miRNA–mRNA pairs, *bna*-miR396 and its putative targets *BnaA06g21030D* and *BnaA01g33410D* were shared by seed samples harvested from both the CQ and YN locations. MiR396 exerts a strong influence in plant development by regulating complex traits; for example, miR396 regulates the transcript levels of the *GRF* group to modulate cell proliferation and elongation (Ercoli et al., 2016), grain size and yield (Che et al., 2015; Li et al., 2016; Chen X. L. et al., 2019), and somatic embryogenesis (Szczygieł-Sommer and Gaj, 2019). Moreover, a loss of function in miR396ef resulted in higher grain size and altered plant architecture in rice (Miao et al., 2020), while over-expression of *oa*-miR396c in rice reduced salt and alkali stress tolerance (Gao et al., 2010). In tomato (*Solanum lycopersicum*), over-expressing a short tandem repeat mimic for miR396a (*STTM396a/396a-88*) resulted in earlier flowering and bigger fruits (Cao et al., 2016). In Arabidopsis, the peanut witches' broom effector PHYLLODY SYMPTOMS 1 (PHYL1) interferes with miR396-mediated regulation of *SHORT VEGETATIVE PHASE* (*SVP*) transcript levels to control flower formation (Yang et al., 2015).

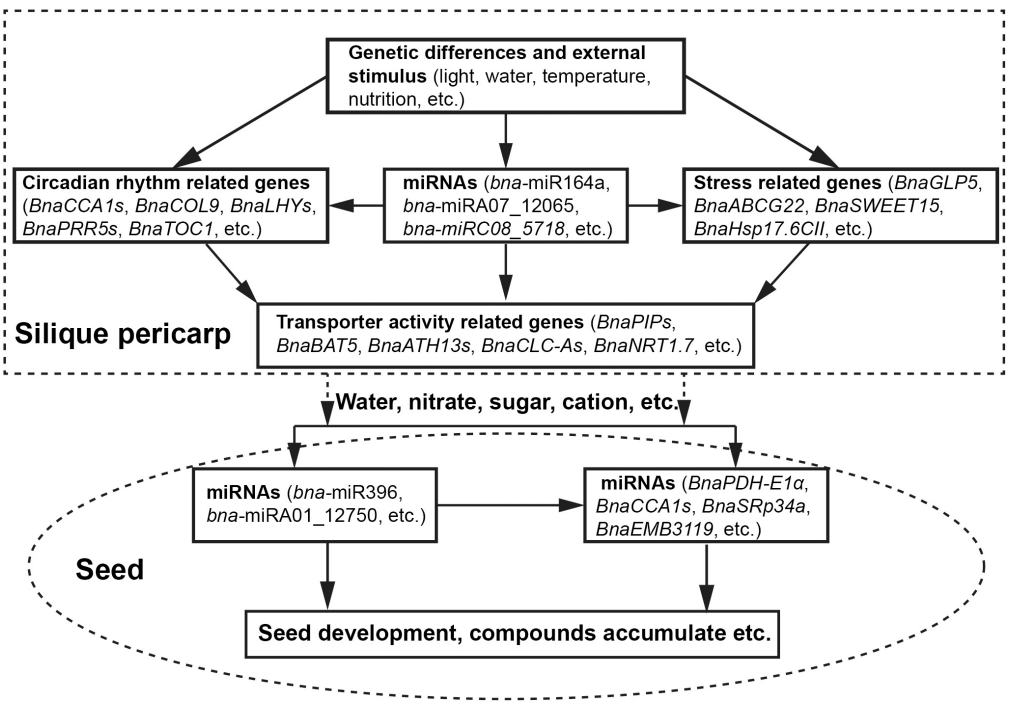


FIGURE 6 | A proposed model for HI determination during the seed-filling stage.

In this study, we also identified the bna-miR396–*BnaA06g21030D*/*BnaA01g33410D* pair. BnaA06g21030D (SER/ARG-rich protein 34A, SRp34a) encodes a member of the Ser-Arg-rich (SR) protein family, which plays multiple roles in post-transcriptional regulation of gene expression by alternative splicing (Sanford et al., 2004; Richardson et al., 2011). BnaA01g33410D is homologous to EMBRYO DEFECTIVE 3119 (EMB3119), which plays an important role in Arabidopsis growth and development (Meinke, 2020). Although many studies have focused on miR396, the roles of the bna-miR396–*Bna.A06SRp34a*/*Bna.A01EMB3119* pair are largely unknown in the context of HI and should be studied in more detail.

Based on our collective results, we propose a potential model for the regulation of HI in *B. napus* during the seed-filling stage (**Figure 6**). Genetic differences and external environmental stimuli affect the expression of circadian rhythm-related genes, stress response genes, and miRNAs. miRNAs, in turn, may modulate circadian rhythm-related genes and stress response genes *via* a transcript cleavage. The resulting modulation of transcript levels of transporter activity-related genes will affect water and solutes flow to the seeds, driving seed development and the accumulation of storage compounds, thus determining seed yield and HI.

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: BIG Data Center under BioProject accession number PRJNA602979.

AUTHOR CONTRIBUTIONS

JL and KL designed the experiments. CZ, WC, XL, BY, LZ, and ZX performed the bioinformatics analysis. BY and LZ performed RNA extractions and qRT-PCR experiments. CZ and WC wrote the manuscript. All the authors reviewed the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31830067 and 31871653), the Talent Project of Chongqing Natural Science Foundation (cstc2021ycjh-bgzxm0033), the 111 Project (B12006), Chongqing Graduate Student Research Innovation Project (CYB20102), Chongqing Special Postdoctoral Science Foundation (cstc2019jcyj-bsh0102), and Germplasm Creation Special Program of Southwest University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.855486/full#supplementary-material

REFERENCES

Aluri, S., and Büttner, M. (2007). Identification and functional expression of the *Arabidopsis thaliana* vacuolar glucose transporter 1 and its role in seed germination and flowering. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2537–2542. doi: 10.1073/pnas.0610278104

Amanullah, and Shah, P. (2010). Timing and rate of nitrogen application influence grain quality and yield in maize planted at high and low densities. *J. Sci. Food Agric.* 90, 21–29. doi: 10.1002/jsfa.3769

Austin, R., Bingham, J., Blackwell, R., Evans, L., Ford, M., Morgan, C., et al. (1980). Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *J. Agric. Sci.* 94, 675–689. doi: 10.1017/S0021859600028665

Basu, U., Bajaj, D., Sharma, A., Malik, N., Daware, A., Narnoliya, L., et al. (2019). Genetic dissection of photosynthetic efficiency traits for enhancing seed yield in chickpea. *Plant Cell Environ.* 42, 158–173. doi: 10.1111/pce.13319

Beche, E., Benin, G., Da Silva, C. L., Munaro, L. B., and Marchese, J. A. (2014). Genetic gain in yield and changes associated with physiological traits in Brazilian wheat during the 20th century. *Eur. J. Agron.* 61, 49–59. doi: 10.1016/j.eja.2014.08.005

Bennett, E. J., Roberts, J. A., and Wagstaff, C. (2011). The role of the pod in seed development: strategies for manipulating yield. *New Phytol.* 190, 838–853. doi: 10.1111/j.1469-8137.2011.03714.x

Bennett, E. J., Roberts, J. A., and Wagstaff, C. (2012). Manipulating resource allocation in plants. *J. Exp. Bot.* 63, 3391–3400. doi: 10.1093/jxb/err442

Boden, S. A., Cavanagh, C., Cullis, B. R., Ramm, K., Greenwood, J., Finnegan, E. J., et al. (2015). *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nat. Plants* 1:14016. doi: 10.1038/nplants.2014.16

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Budak, H., and Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Funct. Integr. Genomics* 15, 523–531. doi: 10.1007/s10142-015-0451-2

Bustin, S. A., Beaulieu, J. F., Huggett, J., Jaggi, R., Kibenge, F. S., Olsvik, P. A., et al. (2010). MIQE précis: practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Mol. Biol.* 11:74. doi: 10.1186/1471-2199-11-74

Büttner, M. (2007). The monosaccharide transporter (-like) gene family in *Arabidopsis*. *FEBS Lett.* 581, 2318–2324. doi: 10.1016/j.febslet.2007.03.016

Byrt, C. S., Zhao, M. C., Kourghi, M., Bose, J., Henderson, S. W., Qiu, J. E., et al. (2017). Non-selective cation channel activity of aquaporin AtPIP2;1 regulated by Ca^{2+} and pH. *Plant Cell Environ.* 40, 802–815. doi: 10.1111/pce.12832

Cao, D. Y., Wang, J., Ju, Z., Liu, Q. Q., Li, S., Tian, H. Q., et al. (2016). Regulations on growth and development in tomato cotyledon, flower and fruit via destruction of miR396 with short tandem target mimic. *Plant Sci.* 247, 1–12. doi: 10.1016/j.plantsci.2016.02.012

Chao, H. B., Raboanatahiry, N., Wang, X. D., Zhao, W. G., Chen, L., Guo, L. X., et al. (2019). Genetic dissection of harvest index and related traits through genome-wide quantitative trait locus mapping in *Brassica napus* L. *Breed. Sci.* 69, 104–116. doi: 10.1270/jsbbs.18115

Che, R. H., Tong, H. N., Shi, B. H., Liu, Y. Q., Fang, S. R., Liu, D. P., et al. (2015). Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat. Plants* 2:15195. doi: 10.1038/nplants.2015.195

Chen, H., Yang, Q., Chen, K., Zhao, S., Zhang, C., Pan, R., et al. (2019). Integrated microRNA and transcriptome profiling reveals a miRNA-mediated regulatory network of embryo abortion under calcium deficiency in peanut (*Arachis hypogaea* L.). *BMC Genomics* 20:392. doi: 10.1186/s12864-019-5770-6

Chen, L. Q., Lin, I. W., Qu, X. Q., Sosso, D., McFarlane, H. E., Londoño, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27, 607–619. doi: 10.3410/f.725398372.793505270

Chen, X. L., Jiang, L. R., Zheng, J. S., Chen, F. Y., Wang, T. S., Wang, M. L., et al. (2019). A missense mutation in Large Grain Size 1 increases grain size and

enhances cold tolerance in rice. *J. Exp. Bot.* 70, 3851–3866. doi: 10.1093/jxb/erz192

Chen, Z. Y., Huo, Q., Yang, H., Jian, H. J., Qu, C. M., Lu, K., et al. (2019). Joint RNA-seq and miRNA profiling analyses to reveal molecular mechanisms in regulating thickness of pod canopy in *Brassica napus*. *Genes* 10:591. doi: 10.3390/genes10080591

D'Ario, M., Griffiths-Jones, S., and Kim, M. S. (2017). Small RNAs: big impact on plant development. *Trends Plant Sci.* 22, 1056–1068. doi: 10.1016/j.tplants.2017.09.009

De Angeli, A., Monachello, D., Ephritikhine, G., Frachisse, J. M., Thomine, S., Gambale, F., et al. (2006). The nitrate/proton antiporter *AtCLCa* mediates nitrate accumulation in plant vacuoles. *Nature* 442, 939–942. doi: 10.3410/f.1040148.489068

Demes, E., Besse, L., Cubero-Font, P., Satiat-Jeunemaitre, B., Thomine, S., and De Angeli, A. (2020). Dynamic measurement of cytosolic pH and $[NO^{3-}]$ uncovers the role of the vacuolar transporter *AtCLCa* in cytosolic pH homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 117, 15343–15353. doi: 10.1101/716050

Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (*Brassica napus L.*): a review. *Field Crops Res.* 67, 35–49. doi: 10.1016/s0378-4290(00)00082-4

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Donald, C. M. (1962). In search of yield. *Aust. Inst. Agric. Sci.* 28, 171–178.

Eom, J. S., Chen, L. Q., Sosso, D., Julius, B. T., Lin, I. W., Qu, X. Q., et al. (2015). SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr. Opin. Plant Biol.* 25, 53–62. doi: 10.1016/j.pbi.2015.04.005

Ercoli, M. F., Rojas, A. M. L., Debernardi, J. M., Palatnik, J. F., and Rodriguez, R. E. (2016). Control of cell proliferation and elongation by miR396. *Plant Signal Behav.* 11:e1184809. doi: 10.1080/15592324.2016.1184809

Fan, S. C., Lin, C. S., Hsu, P. K., Lin, S. H., and Tsay, Y. F. (2009). The *Arabidopsis* nitrate transporter *NRT1.7*, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. *Plant Cell* 21, 2750–2761. doi: 10.3410/f.1165218.628157

Fang, Y. J., Xie, K. B., and Xiong, L. Z. (2014). Conserved miR164-targeted NAC genes negatively regulate drought resistance in rice. *J. Exp. Bot.* 65, 2119–2135. doi: 10.1093/jxb/eru072

Flis, A., Mengin, V., Ivakov, A. A., Mugford, S. T., Hubberten, H. M., Encke, B., et al. (2019). Multiple circadian clock outputs regulate diel turnover of carbon and nitrogen reserves. *Plant Cell Environ.* 42, 549–573. doi: 10.1111/pce.13440

Foulkes, M. J., Reynolds, M. P., and Sylvester-Bradley, R. (2009). Genetic improvement of grain crops-chapter 15: yield potential. *Crop Physiol.* 68, 355–385. doi: 10.1016/j.fcr.2012.05.014

Fu, T. D., and Zhou, Y. M. (2013). Progress and future development of hybrid rapeseed in China. *Eng. Sci.* 11, 13–18.

Fung-Uceda, J., Lee, K., Seo, P. J., Polyn, S., De Veylder, L., and Mas, P. (2018). The circadian clock sets the time of DNA replication licensing to regulate growth in *Arabidopsis*. *Dev. Cell* 45, 101–113.e4. doi: 10.3410/f.732902266.793569724

Gajića, B., Kresovićb, B., Tapanarovaa, A., Životić, L., and Todorovićc, M. (2018). Effect of irrigation regime on yield, harvest index and water productivity of soybean grown under different precipitation conditions in a temperate environment. *Agric. Water Manage.* 210, 224–231. doi: 10.1016/j.agwat.2018.08.002

Gao, P., Bai, X., Yang, L., Lv, D. K., Li, Y., Cai, H., et al. (2010). Over-expression of osa-MIR396c decreases salt and alkali stress tolerance. *Planta* 231, 991–1001. doi: 10.1007/s00425-010-1104-2

Giorgia, L., Irene, L. G., Meurer, J., Jürgen, S., and Serena, S. (2011). The Cytosolic Kinases STY8, STY17, and STY46 Are Involved in Chloroplast Differentiation in *Arabidopsis*. *Plant Physiol.* 157, 70–85. doi: 10.1104/pp.111.182774

Guo, C. K., Xu, Y. M., Shi, M., Lai, Y. M., Wu, X., Wang, H. S., et al. (2017). Repression of miR156 by miR159 regulates the timing of the juvenile-to-adult transition in *Arabidopsis*. *Plant Cell* 29, 1293–1304. doi: 10.3410/f.727650668.793535158

Ham, B. K., Li, G., Kang, B. H., Zeng, F., and Lucas, W. J. (2012). Overexpression of *Arabidopsis* plasmodesmata germin-like proteins disrupts root growth and development. *Plant Cell* 24, 3630–3648. doi: 10.1105/tpc.112.101063

Hay, R. K. M. (1995). Harvest index: a review of its use in plant breeding and crop physiology. *Ann. Appl. Biol.* 126, 197–216. doi: 10.1111/j.1744-7348.1995.tb05015.x

Hirner, B., Fischer, W. N., Rentsch, D., Kwart, M., and Frommer, W. B. (1998). Developmental control of H+ / amino acid permease gene expression during seed development of *Arabidopsis*. *Plant J.* 14, 535–544. doi: 10.1046/j.1365-313x.1998.00151.x

Huang, T. B., López-Giráldez, F., Townsend, J. P., and Irish, V. F. (2012). RBE controls microRNA164 expression to effect floral organogenesis. *Development* 139, 2161–2169. doi: 10.1242/dev.075069

Hütsch, B. W., and Schubert, S. (2017). Harvest index of maize (*Zea mays* L.): are there possibilities for improvement. *Adv. Agron.* 146, 37–82. doi: 10.1016/bs.agron.2017.07.004

Jian, H. J., Lu, K., Yang, B., Wang, T. Y., Zhang, L., Zhang, A. X., et al. (2016). Genome-Wide analysis and expression profiling of the *SUC* and *SWEET* gene families of sucrose transporters in oilseed rape (*Brassica napus* L.). *Front. Plant Sci.* 7:1464. doi: 10.3389/fpls.2016.01464

Kim, S. C., Nusinow, D. A., Sorkin, M. L., Pruneda-Paz, J., and Wang, X. M. (2019). Interaction and regulation between lipid mediator phosphatidic acid and circadian clock regulators. *Plant Cell* 31, 399–416. doi: 10.3410/f.734916270.793556708

Klepek, Y. S., Geiger, D., Stadler, R., Klebl, F., Landouar-Arsivaud, L., Lemoine, R., et al. (2005). *Arabidopsis* POLYOL TRANSPORTER5, a new member of the monosaccharide transporter-like superfamily, mediates H^+-symport of numerous substrates, including myo-inositol, glycerol, and ribose. *Plant Cell* 17, 204–218. doi: 10.1105/tpc.104.026641

Klepek, Y. S., Volke, M., Konrad, K. R., Wippel, K., Hoth, S., Hedrich, R., et al. (2010). *Arabidopsis thaliana* POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H^+ symporters in pollen and young xylem cells. *J. Exp. Bot.* 61, 537–550. doi: 10.3410/f.1430961.908059

Kourghi, M., Nourmohammadi, S., Pei, J. V., Qiu, J. E., McGaughey, S., Tyerman, S. D., et al. (2017). Divalent cations regulate the ion conductance properties of diverse classes of aquaporins. *Int. J. Mol. Sci.* 18:2323. doi: 10.3390/ijms18112323

Kuromori, T., Sugimoto, E., and Shinozaki, K. (2011). *Arabidopsis* mutants of *AtABCG22*, an ABC transporter gene, increase water transpiration and drought susceptibility. *Plant J.* 67, 885–894. doi: 10.1111/j.1365-313x.2011.04641.x

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.3410/f.1156941.617005

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

LeClere, S., Rampey, R. A., and Bartel, B. (2004). *IAR4*, a gene required for auxin conjugate sensitivity in *Arabidopsis*, encodes a pyruvate dehydrogenase E1alpha homolog. *Plant Physiol.* 135, 989–999. doi: 10.1104/pp.104.040519

Legnaioli, T., Cuevas, J., and Mas, P. (2009). *TOC1* functions as a molecular switch connecting the circadian clock with plant responses to drought. *EMBO J.* 28, 3745–3757. doi: 10.3410/f.1166599.627609

Leng, S. H., Zhu, G. R., Deng, X. L., and Jiang, S. (1992). Studies on the sources of the dry matter in the seed of rapeseed. *Acta Agron. Sin.* 04, 250–257.

Li, N., Song, D. J., Peng, W., Zhan, J. P., Shi, J. Q., Wang, X. F., et al. (2019). Maternal control of seed weight in rapeseed (*Brassica napus* L.): the causal link between the size of pod (mother, source) and seed (offspring, sink). *Plant Biotechnol. J.* 17, 736–749. doi: 10.1111/pbi.13011

Li, Q. Q., Wu, G. X., Zhao, Y. P., Wang, B. B., Zhao, B. B., Kong, D. X., et al. (2020). CRISPRCas9-mediated knockout and overexpression studies reveal a role of maize phytochrome C in regulating flowering time and plant height. *Plant Biotechnol. J.* 18, 2520–2532. doi: 10.1111/pbi.13429

Li, S. C., Gao, F. Y., Xie, K. L., Zeng, X. H., Cao, Y., Zeng, J., et al. (2016). The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14, 2134–2146. doi: 10.1111/pbi.12569

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, C., Song, G. Y., Zhou, Y. H., Qu, X. F., Guo, Z. B., Liu, Z. W., et al. (2015). *OsPRR37* and *Ghd7* are the major genes for general combining ability of DTH, PH and SPP in rice. *Sci. Rep.* 5:12803. doi: 10.1038/srep12803

Liu, H., Gu, F. W., Dong, S. Y., Liu, W., Wang, H., Chen, Z. Q., et al. (2016). CONSTANS-like 9 (COL9) delays the flowering time in *Oryza sativa* by

repressing the Ehd1 pathway. *Biochem. Biophys. Res. Commun.* 479, 173–178. doi: 10.1016/j.bbrc.2016.09.013

Liu, W. W., Sun, Q., Wang, K., Du, Q. G., and Li, W. X. (2017). Nitrogen limitation adaptation (NLA) is involved in source-to-sink remobilization of nitrate by mediating the degradation of NRT1.7 in *Arabidopsis*. *New Phytol.* 214, 734–744. doi: 10.1111/nph.14396

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

Lu, K., Chai, Y. R., Zhang, K., Wang, R., Chen, L., Lei, B., et al. (2008). Cloning and characterization of phosphorus starvation inducible *Brassica napus* PURPLE ACID PHOSPHATASE 12 gene family, and imprinting of a recently evolved MITE-minisatellite twin structure. *Theor. Appl. Genet.* 117, 963–975. doi: 10.1007/s00122-008-0836-x

Lu, K., Li, T., He, J., Chang, W., Zhang, R., Liu, M., et al. (2018). qPrimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. *Nucleic Acids Res.* 46, D1229–D1236. doi: 10.1093/nar/gkx725

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., et al. (2017). Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. *Front. Plant Sci.* 8:206. doi: 10.3389/fpls.2017.00206

Lu, K., Xiao, Z. C., Jian, H. J., Peng, L., Qu, C. M., Fu, M. L., et al. (2016). A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in *Brassica napus*. *Sci. Rep.* 6:36452. doi: 10.1038/srep36452

Luo, X., Ma, C. Z., Yue, Y., Hu, K. N., Li, Y. Y., Duan, Z. Q., et al. (2015). Unravelling the complex trait of harvest index in rapeseed (*Brassica napus* L.) with association mapping. *BMC Genomics* 16:379. doi: 10.1186/s12864-015-1607-0

Macho-Rivero, M. A., Herrera-Rodríguez, M. B., Brejcha, R., Schäffner, A. R., Tanaka, N., Fujiwara, T., et al. (2018). Boron toxicity reduces water transport from root to shoot in *Arabidopsis* plants. Evidence for a reduced transpiration rate and expression of major pip aquaporin genes. *Plant Cell Physiol.* 59, 836–844. doi: 10.1093/pcp/pcy026

Maere, S., Heymans, K., and Kuiper, M. J. B. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Manara, A., DalCorso, G., Guzzo, F., and Furini, A. (2015). Loss of the atypical kinases ABC1K7 and ABC1K8 changes the lipid composition of the chloroplast membrane. *Plant Cell Physiol.* 56, 1193–1204. doi: 10.1093/pcp/pcv046

Manara, A., Dalcorso, G., Leister, D., Jahns, P., Baldan, B., and Furini, A. (2014). *AtSIA1* and *AtOSA1*: two ABC1 proteins involved in oxidative stress responses and iron distribution within chloroplasts. *New Phytol.* 201, 452–465. doi: 10.1111/nph.12533

Mcclung, C. R., and Gutiérrez, R. A. (2010). Network news: prime time for systems biology of the plant circadian clock. *Curr. Opin. Genet. Dev.* 20, 588–598. doi: 10.1016/j.gde.2010.08.010

Meinke, D. W. (2020). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in *Arabidopsis*. *New Phytol.* 226, 306–325. doi: 10.1111/nph.16071

Mentzen, W. I., Peng, J. L., Ransom, N., Nikolau, B. J., and Wurtele, E. S. (2008). Articulation of three core metabolic processes in *Arabidopsis*: fatty acid biosynthesis, leucine catabolism and starch metabolism. *BMC Plant Biol.* 8:76. doi: 10.1186/1471-2229-8-76

Miao, C. B., Wang, D., He, R. Q., Liu, S. K., and Zhu, J. K. (2020). Mutations in MIR396e and MIR396f increase grain size and modulate shoot architecture in rice. *Plant Biotechnol. J.* 18, 491–501. doi: 10.1111/pbi.13214

Nakamichi, N., Kita, M., Niinuma, K., Ito, S., Yamashino, T., Mizoguchi, T., et al. (2007). *Arabidopsis* clock-associated Pseudo-Response Regulators PRR9, PRR7 and PRR5 coordinately and positively regulate flowering time through the canonical CONSTANS-dependent photoperiodic pathway. *Plant Cell Physiol.* 48, 822–832. doi: 10.1093/pcp/pcm056

Nakamichi, N., Kudo, T., Makita, N., Kiba, T., Kinoshita, T., and Sakakibara, H. (2020). Flowering time control in rice by introducing *Arabidopsis* clock-associated PSEUDO-RESPONSE REGULATOR 5. *Biosci. Biotechnol. Biochem.* 84, 970–979. doi: 10.1080/09168451.2020.1719822

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.3410/f.727375548.793529669

Prasad, P. V. V., Boote, K. J., Allen, L. H., Sheehy, J. E., and Thomas, J. M. G. (2006). Species, ecotype and cultivar differences in spikelet fertility and harvest index of rice in response to high temperature stress. *Field Crops Res.* 95, 398–411. doi: 10.1016/j.fcr.2005.04.008

Qu, C. M., Fu, F. Y., Liu, M., Zhao, H. Y., Liu, C., Li, J. N., et al. (2015). Comparative transcriptome analysis of recessive male sterility (RGMS) in sterile and fertile *Brassica napus* lines. *PLoS One* 10:e0144118. doi: 10.1371/journal.pone.0144118

Qu, C., Zhao, H., Fu, F., 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.* 7:1755. doi: 10.3389/fpls.2016.01755

Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P., and Theres, K. (2008). Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in *Arabidopsis thaliana*. *Plant J.* 55, 65–76. doi: 10.1111/j.1365-313x.2008.03483.x

Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., et al. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS One* 6:e24542. doi: 10.1371/journal.pone.0024542

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sadras, V. O., and Lawson, C. (2011). Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of south Australian wheat varieties released between 1958 and 2007. *Crop Pasture Sci.* 62, 533–549. doi: 10.1071/cp11060

Sanford, J. R., Gray, N. K., Beckmann, K., and Cáceres, J. F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* 18, 755–768. doi: 10.1101/gad.286404

Sawada, Y., Toyooka, K., Kuwahara, A., Sskata, A., Nagano, M., Satio, K., et al. (2009). *Arabidopsis* bile acid: sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. *Plant Cell Physiol.* 50, 1579–1586. doi: 10.1093/pcp/pcp110

Seo, P. J., Park, J. M., Kang, S. K., Kim, S. G., and Park, C. M. (2011). An *Arabidopsis* senescence-associated protein SAG29 regulates cell viability under high salinity. *Planta* 233, 189–200. doi: 10.1007/s00425-010-1293-8

Shan, J. W., Song, W., Zhou, J., Wang, X. H., Xie, C. H., Gao, X. X., et al. (2013). Transcriptome analysis reveals novel genes potentially involved in photoperiodic tuberization in potato. *Genomics* 102, 388–396. doi: 10.1016/j.ygeno.2013.07.001

Shen, J. X., Fu, T. D., Yang, G. S., Ma, C. Z., and Tu, J. X. (2010). Genetic analysis of rapeseed self-incompatibility lines reveals significant heterosis of different patterns for yield and oil content traits. *Plant Breed.* 124, 111–116. doi: 10.1111/j.1439-0523.2004.01069.x

Sommer, A., Geist, B., Da Ines, O., Gehwolf, R., Schäffner, A. R., and Obermeyer, G. (2008). Ectopic expression of *Arabidopsis thaliana* plasma membrane intrinsic protein 2 aquaporins in lily pollen increases the plasma membrane water permeability of grain but not of tube protoplasts. *New Phytol.* 180, 787–797. doi: 10.1111/j.1469-8137.2008.02607.x

Sun, H. M., Qian, Q., Wu, K., Luo, J. J., Wang, S. S., Zhang, C. W., et al. (2014). Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Genet.* 46, 652–656. doi: 10.1038/ng.2958

Szczygieł-Sommer, A., and Gaj, M. D. (2019). The miR396-GRF regulatory module controls the embryogenic response in *Arabidopsis* via an auxin-related pathway. *Int. J. Mol. Sci.* 20:5221. doi: 10.3390/ijms20205221

Tang, J. Y., and Chu, C. C. (2017). MicroRNAs in crop improvement: fine-tuners for complex traits. *Nat. Plants* 3:17077. doi: 10.1038/nplants.2017.77

Tang, Y. Y., Liu, H. H., Guo, S. Y., Wang, B., Li, Z. T., Chong, K., et al. (2018). OsmiR396d miRNA affects gibberellin and brassinosteroid signaling to regulate plant architecture. *Plant Physiol.* 176, 946–959. doi: 10.1104/pp.17.00964

Tian, T., Liu, Y., Yan, H. Y., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Tong, W., Imai, A., Tabata, R., Shigenobu, S., Yamaguchi, K., Yamada, M., et al. (2016). Polyamine resistance is increased by mutations in a nitrate transporter

gene *NRT1.3* (*AtNPF6.4*) in *Arabidopsis thaliana*. *Front. Plant Sci.* 7:834. doi: 10.3389/fpls.2016.00834

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, 562–578. doi: 10.3410/f.14267340.15779565

Unkovich, M., Baldock, J., and Forbes, M. (2010). Variability in harvest index of grain crops and potential significance for carbon accounting: examples from Australian agriculture. *Adv. Agron.* 105, 173–218. doi: 10.1016/s0065-2113(10)05005-4

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046

Wang, C. Y., Zhang, S. C., Yu, Y., Luo, Y. C., Liu, Q., Ju, C. L., et al. (2014). MiR397b regulates both lignin content and seed number in *Arabidopsis* via modulating a laccase involved in lignin biosynthesis. *Plant Biotechnol. J.* 12, 1132–1142. doi: 10.1111/pbi.12222

Wang, H., Zhang, L. Y., Tao, Y., Wang, Z. D., Shen, D., and Dong, H. S. (2020). Transmembrane helices 2 and 3 determine the localization of plasma membrane intrinsic proteins in eukaryotic cells. *Front. Plant Sci.* 10:1671. doi: 10.3389/fpls.2019.01671

Wang, X. D., Zheng, M., Liu, H. F., Zhang, L., Chen, F., Zhang, W., et al. (2020). Fine-mapping and transcriptome analysis of a candidate gene controlling plant height in *Brassica napus* L. *Biotechnol. Biofuels* 13:42.

Xu, H., Li, X. F., Zhang, H., Wang, L. C., Zhu, Z. G., Gao, J. P., et al. (2020). High temperature inhibits the accumulation of storage materials by inducing alternative splicing of *OsbZIP58* during filling stage in rice. *Plant Cell Environ.* 43, 1879–1896. doi: 10.1111/pce.13779

Yang, C. Y., Huang, Y. H., Lin, C. P., Lin, Y. Y., Hsu, H. C., Wang, C. N., et al. (2015). MicroRNA396-targeted *SHORT VEGETATIVE PHASE* is required to repress flowering and is related to the development of abnormal flower symptoms by the phyllody symptoms1 effector. *Plant Physiol.* 168, 1702–1716. doi: 10.1104/pp.15.00307

Yang, M., He, J. B., Wan, S. B., Li, W. Y., Chen, W. J., Wang, Y. M., et al. (2021). Fine mapping of the *BnaC04.BIL1* gene controlling plant height in *Brassica napus* L. *BMC Plant Biol.* 21:359. doi: 10.1186/s12870-021-03137-9

Yang, X. H., Xia, X. Z., Zhang, Z. Q., Nong, B. X., Zeng, Y., Xiong, F. Q., et al. (2017). QTL mapping by whole genome re-sequencing and analysis of candidate genes for nitrogen use efficiency in rice. *Front. Plant Sci.* 8:1634. doi: 10.3389/fpls.2017.01634

Yang, X. Z., and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27, 2614–2615. doi: 10.1093/bioinformatics/btr430

Zhang, L. Y., Zhang, C., Yang, B., Xiao, Z. C., Ma, J. Q., Liu, J. S., et al. (2020). Genome-wide identification and expression profiling of monosaccharide transporter genes associated with high harvest index values in rapeseed (*Brassica napus* L.). *Genes* 11:653. doi: 10.3390/genes11060653

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 02 August 2022
DOI 10.3389/fpls.2022.927418

OPEN ACCESS

EDITED BY
Sarvajeet Singh Gill,
Maharshi Dayanand University, India

REVIEWED BY
Gurpreet Kaur,
Punjab Agricultural University, India
Dinggang Zhou,
Hunan University of Science and Technology, China

*CORRESPONDENCE
Ruijun Duan
ruijunduan@163.com

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

RECEIVED 24 April 2022
ACCEPTED 01 July 2022
PUBLISHED 02 August 2022

CITATION
Xiong H, Wang R, Jia X, Sun H and Duan R (2022) Transcriptomic analysis of rapeseed (*Brassica napus*. L.) seed development in Xiangride, Qinghai Plateau, reveals how its special eco-environment results in high yield in high-altitude areas. *Front. Plant Sci.* 13:927418. doi: 10.3389/fpls.2022.927418

Transcriptomic analysis of rapeseed (*Brassica napus*. L.) seed development in Xiangride, Qinghai Plateau, reveals how its special eco-environment results in high yield in high-altitude areas

Huiyan Xiong[1], Ruisheng Wang[2], Xianqing Jia[3], Hezhe Sun[1] and Ruijun Duan[4]*

[1]College of Agriculture and Animal Husbandry, Qinghai University, Xining, China, [2]Academy of Agricultural and Forestry Sciences of Qinghai University, Key Laboratory of Spring Rape Genetic Improvement of Qinghai Province, Rapeseed Research and Development Center of Qinghai Province, Xining, China, [3]Key Laboratory of Plant Nutrition and Fertilizer, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China, [4]College of Eco-Environmental Engineering, Qinghai University, Xining, China

As one of the most important oil crops, rapeseed (*Brassica napus*) is cultivated worldwide to produce vegetable oil, animal feed, and biodiesel. As the population grows and the need for renewable energy increases, the breeding and cultivation of high-yield rapeseed varieties have become top priorities. The formation of a high rapeseed yield is so complex because it is influenced not only by genetic mechanisms but also by many environmental conditions, such as climatic conditions and different farming practices. Interestingly, many high-yield areas are located in special eco-environments, for example, in the high-altitude Xiangride area of the Qinghai Plateau. However, the molecular mechanisms underlying the formation of high yields in such a special eco-environment area remain largely unknown. Here, we conducted field yield analysis and transcriptome analysis in the Xiangride area. Compared with the yield and environmental factors in the Xinning area (a low-yielding area), we found that the relatively longer daylight length is the key to high rapeseed yield in the Xiangride area, which leads up to a 52.1% increase in rapeseed yield, especially the increase in thousand seed weight and silique number (SN). Combined with transcriptome H-cluster analysis and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional analyses, we can assume that the grain development of rapeseed in the Xiangride area is ahead of schedule and lasts for a long time, leading to the high-yield results in the Xiangride area, confirmed by the expression analysis by quantitative real-time polymerase chain reaction (qRT-PCR) of yield-related genes. Our results provide valuable information for further exploring the molecular mechanism

underlying high yield in special ecological environments and provide a helpful reference for studying seed development characteristics in special-producing regions for *Brassica napus*.

KEYWORDS

rapeseed (*Brassica napus* L.), special environment, plateau, daylight length, transcriptome analysis, seed development

Introduction

Rapeseed (*Brassica napus*) is one of the most important oil crops and is cultivated worldwide to produce vegetable oil, animal feed, biodiesel, etc. (Chalhoub et al., 2014; Mason and Snowdon, 2016; Hu et al., 2021). With rapid increases in population and the need for renewable energy, the demand for rapeseed also continually rises. Therefore, the breeding and cultivation of high-yield rapeseed varieties have become top priorities (Slafer, 2003; Hu et al., 2009; Khan et al., 2021). Seed yield can be improved by taking the direct component traits and the other indirect contributing traits into consideration. For the representative oil crop rapeseed, high yield is determined mainly by three yield component traits: silique number (SN), seed number per silique (SPS), and seed weight/thousand seed weight (SW/TSW). SN and SPS determine the total number of seeds per plant, and SW/TSW has the most significant effect on seed yield (Leon, 1993; Diepenbrock, 2000; Shi et al., 2009; Ding et al., 2012; Luo et al., 2015, 2017). High rapeseed yield is also closely associated with many yield-related traits, such as plant height (PH) and silique length (SL). In addition to being controlled by yield-related genes, an increasing number of studies have shown that these traits are also highly influenced by environmental conditions (Quijada et al., 2006; Udall et al., 2006; Chen et al., 2007; Li et al., 2014; Fu et al., 2015; Zhao et al., 2016, 2021; Zheng et al., 2017; Raboanatahiry et al., 2018).

The effects of the environmental factors on all growth and development processes occur throughout the whole plant life cycle (Quarrie et al., 2006). For the same cultivars, various planting environments lead to different yield potentials, and some special eco-environments can lead to high yields, for example, Taoyuan Township, Yunnan Province and Xiangride area, Qaidam Basin, Qinghai Province in China (Ying et al., 1998a,b; Katsura et al., 2008). The Qaidam Basin, a vast intermountain basin in the northwest Qinghai Province, Qinghai-Tibetan Plateau, is also the largest basin in China with the highest elevation. The agricultural distribution area of the Qaidam Basin has sparse precipitation, a dry climate, low relative humidity, low moisture content in the air, low humidity, and high atmospheric transparency. Significantly, the annual sunshine hours exceed 3,000, and the yearly total solar radiation is approximately 70. The Xiangride area in the Qaidam Basin has become one of the typical high-yielding areas of crops in China due to its special geographical and ecological environment (Su and Pan, 1981; Shi et al., 2014). However, the genetic and molecular mechanisms underlying the high yield formation of crops in this special high-yield Xiangride area are still unclear.

Seed development is an important stage in determining the yield of rapeseed (Hay and Schwender, 2011; Basnet et al., 2013; Borisjuk et al., 2013; Liu et al., 2015a; Gupta et al., 2017). A previous study suggested that cellular activities during seed filling in *B. napus* begin with sugar mobilization, followed by sequential surges in amino acid, lipid, and storage protein synthesis (Hajduch et al., 2006). Based on three stages in seed filling, they divided the expression trends into different functional stages. The first stage includes proteins expressed mainly at the early stages of seed filling. These proteins are involved in glycolysis, respiration, metabolism of sugars, signal transduction, metabolism of amino acids, proteolysis, and defense. Proteins of the second stage, involved in photosynthesis and lipid metabolism, exhibited the highest expression at the midpoint of seed filling. Finally, detoxification, seed maturation, and seed storage proteins are highly abundant at the end stage of seed filling (Hajduch et al., 2006; Huang et al., 2013). Seed development is highly affected by environmental factors, such as sunshine duration, temperature, and water. Thus, we also proposed that the special eco-environment in the Xiangride area could affect the seed development stages and further cause the high yield of rapeseed.

In this study, we performed yield component analysis and comparative transcriptomic analysis of seed development of Qingza 5, a spring rapeseed variety, in different environments to determine the primary yield component responsible for differences in high-yield environments. The differentially expressed genes (DEGs) related to high yield were analyzed, and the high-yield-related genes were also explored by RNA-Seq and quantitative real-time polymerase chain reaction (qRT-PCR). The research results will help to better understand the molecular mechanism of high yield formation in the Xiangride area of the Qaidam Basin, and have important theoretical and practical

significance for spring rapeseed breeding and improving yield potential.

Materials and methods

Plant materials and growth conditions

Field experiments were conducted in 2015 in the Xining area (XN, 101°49′17″E, 36°34′3″N, altitude 2,320 m) and Xiangride area (XRD, 97°48′9.68″E, 36°4′13.68″N, altitude 2,997 m), Qinghai Province, China, during the rapeseed growing season from April to August. The average annual temperature and rainfall of the XN area were 6.4°C and 306.2 mm, respectively, and those of the XRD area were 4.2°C and 142.5 mm, respectively (for details, see Additional File 1: **Supplementary Table 1**). A widely planted rapeseed (*B. napus*) variety, Qingza 5, was selected and planted at the experimental farm of Qinghai University (Xining, Qinghai Province, China) and Xiangride farm (Dulan, Qinghai Province, China). The field experiment followed a randomized complete block design with three replicates for every 20 plots. Each plot had an acreage of 12 m^2 and consisted of 200 plants. The planting density was 20 cm between plants within each row and 30 cm between rows. Field management followed regular planting practices, and the planting practices in the two places were identical. At the grain filling stage, three replicates of siliques from 30 or 40 days after flowering were collected for RNA extraction. All seeds were gathered, snap-frozen in liquid nitrogen, and kept at −80°C for further use.

Measurement of rapeseed yield traits

Measurements of yield per plant, SN, thousand seed weight (TSW), and SPS were performed after harvest. Ten plants growing uniformly in three replications from each plot in the producing area were chosen for trait evaluation. TSW was calculated based on the average weight of 1,000 fully developed open-pollinated dry seeds.

Collection of climate information

General climate data of XN and XRD in 2015, including average temperature, rainfall, daylight length, and average soil temperature (of 20 cm), were obtained from the China Meteorological Administration.[1]

1 http://www.cma.gov.cn/

RNA extraction, RNA-seq library construction, sequencing, and data analysis

Total RNA was isolated by using the TRIzol kit (Invitrogen, Carlsbad, CA, United States) and purified using an mRNA purification kit (Promega, Shanghai, China) following the manufacturer's instructions with some modifications. RNA degradation and contamination were monitored on 1% agarose gels and treated with RNase-free DNase I (Thermo Fisher Scientific, Waltham, MA, United States) to remove any contaminating DNA. The quality and integrity of the extracted RNAs were assessed using the NanoPhotometer® spectrophotometer (Implen, Westlake Village, CA, United States) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, United States). All these RNA samples should display R260/280 at 1.8–2.0, and the threshold of the RNA integrity number (RIN) was set to at least 8. After the quality assessment, 3 μg of RNA per sample was further processed by the purification of polyA-containing mRNA, mRNA fragmentation, double-stranded cDNA synthesis, and polymerase chain reaction (PCR) amplification, and RNA-Seq libraries were generated using the NEBNext® Ultra RNA Library Prep Kit for Illumina® (NEB, Ipswich, United Kingdom) following the manufacturer's instructions. The final cDNA libraries were sequenced on an Illumina HiSeq 2000 platform by the BGI Tech Solutions Co., Ltd. (BGI-Tech, Shenzhen, Guangdong, China). To preferentially select cDNA fragments of preferentially 150–200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Brea, CA, United States). At the same time, the Q20, Q30, and GC contents of the clean data were calculated.

Rapeseeds "*Darmor-bzh*" reference genome and gene annotation files were downloaded from the genome website BRAD (http://brassicadb.cn/). The index of the reference genome was built using Bowtie v2.2.3, and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12. HTSeq v0.6.1 was used to count the read numbers mapped to each gene. Then, the fragments per kilobase of transcript per million mapped reads (FPKM) value of each gene was calculated based on the length of the gene and read count mapped to this gene. Differential expression analysis of the two conditions/groups was performed using the DESeq R package (1.18.0). The resulting *P*-values were adjusted using Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted *P*-value <0.05 found by DESeq were considered differentially expressed genes. Gene Ontology (GO) enrichment analysis of differentially expressed genes (DEGs) was implemented by the GOseq R package, in which gene length bias was corrected. GO terms with corrected *P* values less than 0.05 were considered significantly enriched DEGs. KOBAS

software was used to test the statistical enrichment of DEGs in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Quantitative real-time PCR analysis

The transcript levels of 11 candidate DEGs of yield-related genes were verified by quantitative real-time PCR (qRT-PCR). Total RNA was treated with DNase, and first-strand cDNA was generated using an AMV First Strand cDNA Synthesis Kit (Sangon, Shanghai, China). SYBR-based qRT-PCRs (SYBR Green I, ABI, Bel Air, MD, United States) were performed on a LightCycler 480 system (Roche, Basel, Switzerland) using the following reaction conditions: 95°C for 3 min followed by 40 cycles of 95°C for 15 s and 60°C for 40 s. The actin gene was used as the internal standard. Three independent biological and technological replicates were performed. The relative transcription level was calculated according to the $2^{-\Delta\Delta Ct}$ method with actin reference genes as a control. Primers are available in **Supplementary Table 2**.

Results

Comparison of yield and its contributing traits between two different areas

To evaluate the differences in rapeseed yield among different producing areas and uncover the effects of unique environmental factors on rapeseed yield, all field experiments in this study were conducted in the Xining area (XN) and Xiangride area (XRD) during the rapeseed growing season from April to August in 2015. XRD is a representative high-altitude and high-yield producing area in the Qinghai Plateau. After harvest, we compared the yield of rapeseed in these two different regions and found that, as expected, the yield per plant was significantly higher in the XRD area. To dissect yield traits, we then analyzed the three main factors of yield formation, including SN, thousand seed weight (TSW), and SPS. The results showed no difference in SPS between the two areas, but the SN and TSW of rapeseed in the XRD area were significantly greater than the SN and SW of rapeseed in the XN area (**Figure 1** and **Supplementary Table 3**). Specifically, the SN and yield per rapeseed plant in the XRD area were 59.6 and 52.1% higher than the SN and yield per rapeseed plant in the XN area.

To reveal the main difference in environmental factors between the two growth areas, we compared the average temperature (AT), rainfall (RF), daylight length (DL), and average soil temperature of 20 cm (AST) recorded in the planting season in 2015 (**Figure 2A**). The AT, RF, and AST were slightly different, but DL was significantly different between the XN and XRD areas. At the same time, compared with XN, DL is the only factor with better data in the XRD area (**Supplementary Table 1**). To assess the relative contributions of these environmental factors, we performed a principal component analysis (PCA) (**Figure 2B**). The top two principal components combined to explain a robust 89.7% of the overall variation and have a high degree of interpretation of all environmental factors (**Supplementary Table 1** and **Supplementary Figure 1**). The results showed that DL is the main factor contributing to the environmental difference between the XN and XRD areas. Taken together, these results indicate that the higher TSW and SN were associated with longer daylight length in the XRD area.

Transcriptome sequencing

In *Brassica napus*, 30–40 days after flowering was the fastest-growing period for seed development, which determined mainly yield formation (Basnet et al., 2013; Liu et al., 2015a). To understand the possible mechanisms of the molecular regulation of higher yields in the special XRD eco-environment, we conducted a comparative transcriptomic study of rapeseed at 30 and 40 days after flowering (DAF) in two different areas (XRD and XN) with three biological replicates. In total, 210.4 million short reads were generated, with 203.9 million high-quality clean reads selected for further analysis. After trimming the adaptor sequences and removing low-quality and short reads, a total of 7.94 G, 7.42 G, 7.98 G, and 7.19 G nucleotides were generated from the XN1 (Xining area, 30 DAF), XN2 (Xining area, 40 DAF), XRD1 (Xiangride area, 30 DAF), and XRD2 (Xiangride area, 40 DAF) libraries, respectively. The minimum Q30 was more than 93%, and the average GC (guanine-cytosine) content was 47.69%, suggesting that the sequencing data were highly accurate and reliable. Then, the total reads of XN1 (53,050,694), XN2 (49,526,474), XRD1 (53,324,798), and XRD2 (48,055,842) were aligned with the reference genome. On average, 76.6%, 81.04%, 80.63%, and 81.14% of the reads were successfully mapped to the reference genome. By comparison with the reference rapeseed genome, all 203.9 million clean reads were assembled into 104,695 genes using Cufflinks, providing massive data for further analysis. Detailed information on quality control and distribution is shown in **Supplementary Tables 4**, **5**. These results suggested that the RNA-Seq data used in the present study were highly reliable.

Identification of differentially expressed genes

To identify the differentially expressed genes (DEGs) between the XN and XRD samples that had relatively high abundance, the fragments per kilobase of exon per million

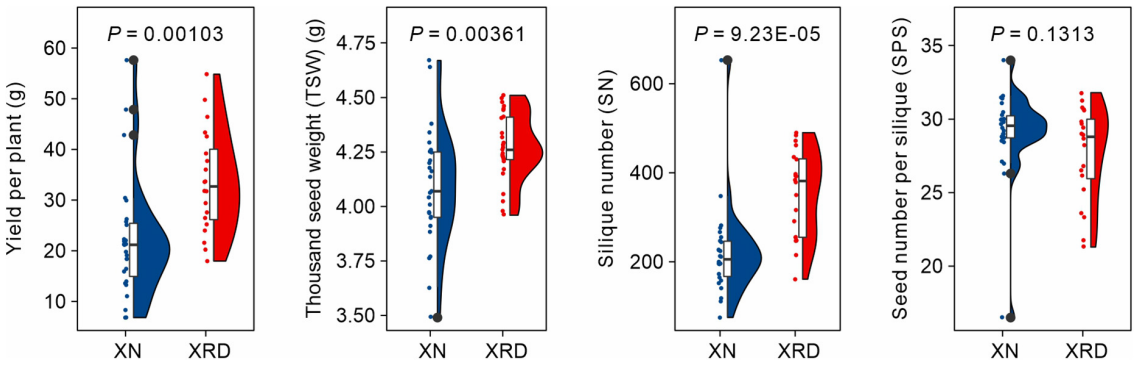


FIGURE 1
Comparisons of yield-related traits of rapeseed plants in XN and XRD areas.

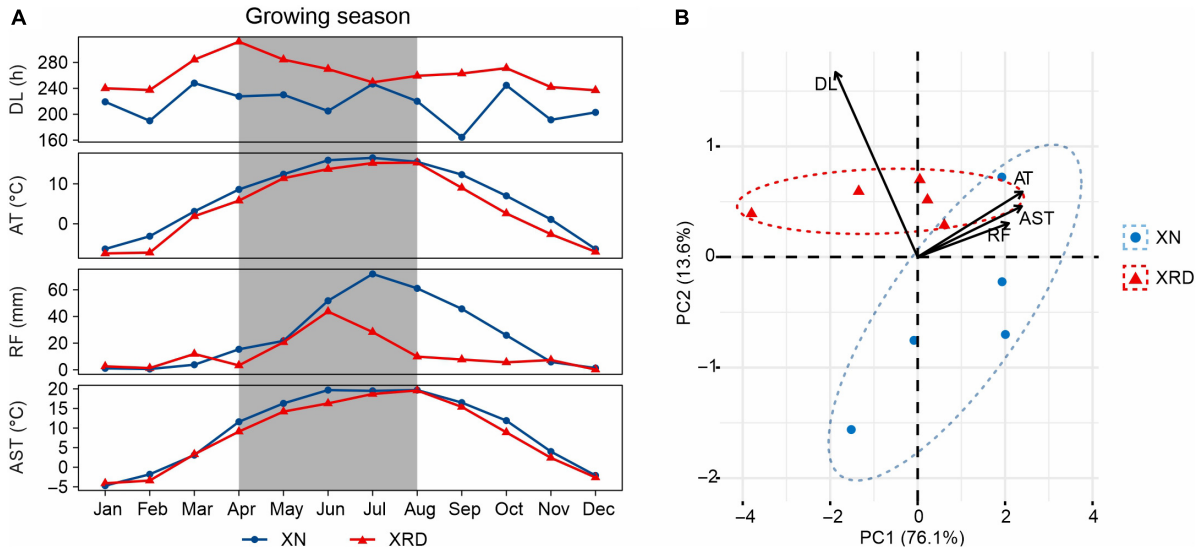


FIGURE 2
Comparison of the environmental factors of the XN and XRD areas. **(A)** General climate information of XN and XRD areas in 2015. Months with the gray background are the growing season of rapeseed (April to August). AT, average temperature; RF, rainfall; DL, daylight length; AST, average soil temperature at 20 cm. All data were obtained from the China Meteorological Administration and are listed in **Supplementary Table 1**. **(B)** Principal component analysis of average temperature, rainfall, daylight length, and average soil temperature of XN and XRD in the growing season. Together, these two axes explain a robust 89.7% of the overall variation.

mapped reads (FPKM) value was used to normalize the gene expression levels. The DEGs were defined as the fold change of the normalized reads per kilobase of transcript per million mapped reads (RPKM) at log2Ratio $\geq$1 and q value $\leq$0.005. Their detailed expression information is shown in **Supplementary Table 6**. Compared with transcriptomic data from XN and XRD areas, the results showed that a total of 2,242 genes were differentially expressed between the two regions at 30 DAF, of which 1,570 genes were downregulated, and 672 genes were upregulated in rapeseeds planted in XRD compared with the rapeseeds planted in the XN region. A total of 366 genes were differentially expressed at 40 days between rapeseeds planted in XN and XRD, of which 231 genes and 135 genes were downregulated and upregulated in rapeseeds planted in XRD, respectively. There were also 2,654 DEGs identified in rapeseeds planted in the XN area at different times, of which 1,941 genes and 713 genes were downregulated and upregulated at 40 days compared with 30 days, respectively. In addition, 164 genes were differentially expressed in rapeseeds planted in the XRD area at different times, of which 80 and 84 genes were downregulated and upregulated, respectively (**Figure 3A**).

To further study the effects of different time points and regions on rapeseed yield, we carried out a Venn diagram of the DEGs (**Figure 3B** and **Supplementary Table 7**). The results showed 115 common DEGs at different times, representing the changes in rapeseed seed development from 30 to

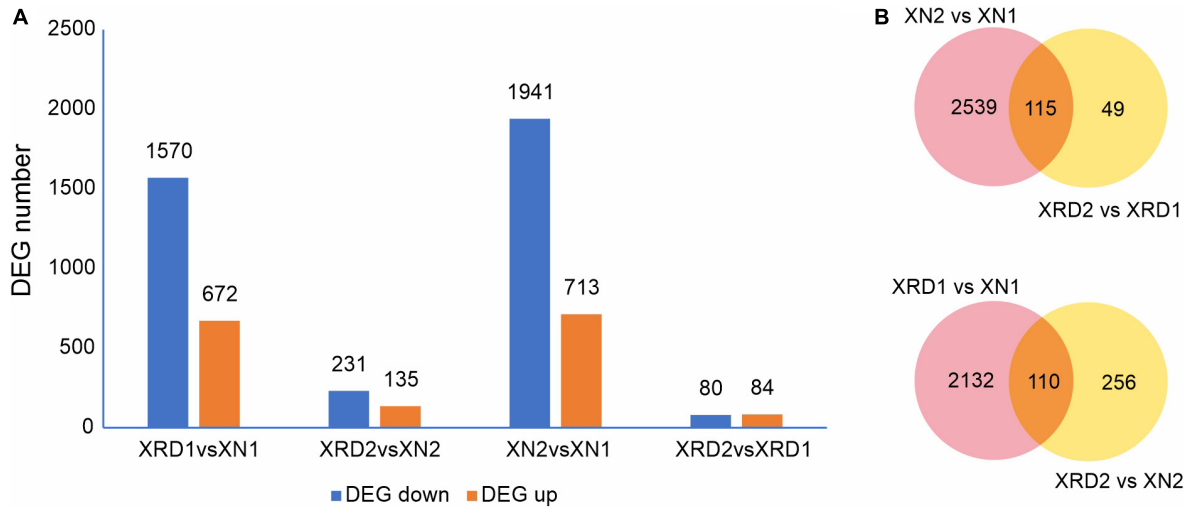


FIGURE 3
DEGs in different regions at different time points. **(A)** Numbers of upregulated and downregulated DEGs in XN and XRD at different time points. **(B)** Venn diagram of DEGs in different regions at different time points.

40 days. Among these DEGs, the BnaC07g42500D (xyloglucan endotransglucosylase/hydrolase protein 24, XTH24) gene was downregulated by more than 100 times in the XN region. In contrast, in the XRD region, this gene was downregulated by more than ten times, indicating that this gene may control the size of the pod by controlling the cell wall decomposition at 30 days. The later disappearance indicated that the pod was transferred into the material accumulation process. The BnaC05g31880D (vacuolar-processing enzyme delta-isozyme) gene has similar trends and may play a similar role in the rapeseed development stage. Most of the upregulated DEGs were dramatically changed in the XN region compared to the downregulated DEGs, but the changes in the XRD region were relatively small, except for BnaC06g13400D and BnaA07g15220D. These two genes are omega-hydroxypalmitate O-feruloyl transferases, which may be involved in the synthesis of aromatics of the suberin polymer, specifically affecting the accumulation of the ferulate constituent of suberin in roots and seeds. The 49 DEGs in **Figure 2B** were unique to the seed development of the special eco-environment XRD region, including 34 upregulated DEGs and 15 downregulated DEGs. Some regulatory genes were found in the unique development-related genes in the XRD area, including auxin-repressed 12.5 kDa protein involved in plant hormone regulatory pathways, adenosine triphosphate (ATP) synthetase involved in energy conversion and substance transfer transporter (organic cation/carnitine transporter), protein kinases (receptor-like serine/threonine-protein kinase), lipid degrading enzyme (probable peroxygenase 3), and superoxide dismutase (SOD) involved in later oxidation, degradation, etc. When DEGs in different regions were merged, 110 DEGs in **Figure 2B** may represent that these genes were related to the differences between the two areas, which can be considered as regional difference DEGs. In these DEGs, circadian clock-related genes (Protein LHY, CCA1), hormone-related genes (Protein EXORDIUM), and other regulatory proteins (TAR1_KLULA Protein), as well as lipid-related enzymes (Delta-9 acyl-lipid desaturase), were common regional DEG.

To further investigate the transcriptomic dynamics at different time points in the XN and XRD areas, we also used the H-cluster (hierarchical clustering analysis) method to cluster the relative expression level of DEGs based on the FPKM values (**Figure 4** and **Supplementary Table 8**). The results of the hierarchical clustering analysis showed that subcluster_1 (1,611 genes), subcluster_5 (681 genes), and subcluster_6 (224 genes) had the same trend of change, and the expression in both regions decreased after 40 days. The changing trend of subcluster_3 (394 genes) and subcluster_4 (246 genes) was the same, and the expression in both regions increased after 40 days. Subcluster_2 (315 genes) was relatively special, and the changes were relatively gentle, especially in the XRD area (**Figure 3A**). From the clustering results, XRD1, XRD2, and XN2 can be seen to have a stronger correlation (**Figure 3B**), indicating that the development of grains in the XRD region occurred significantly earlier and lasted longer than the development of grains in the XN region.

Yield-related gene analysis

A previous study showed that a series of *Brassica napus* homologous genes of *Arabidopsis thaliana* yield-related genes were related to nine traits, such as SY (seed yield), SW (seed

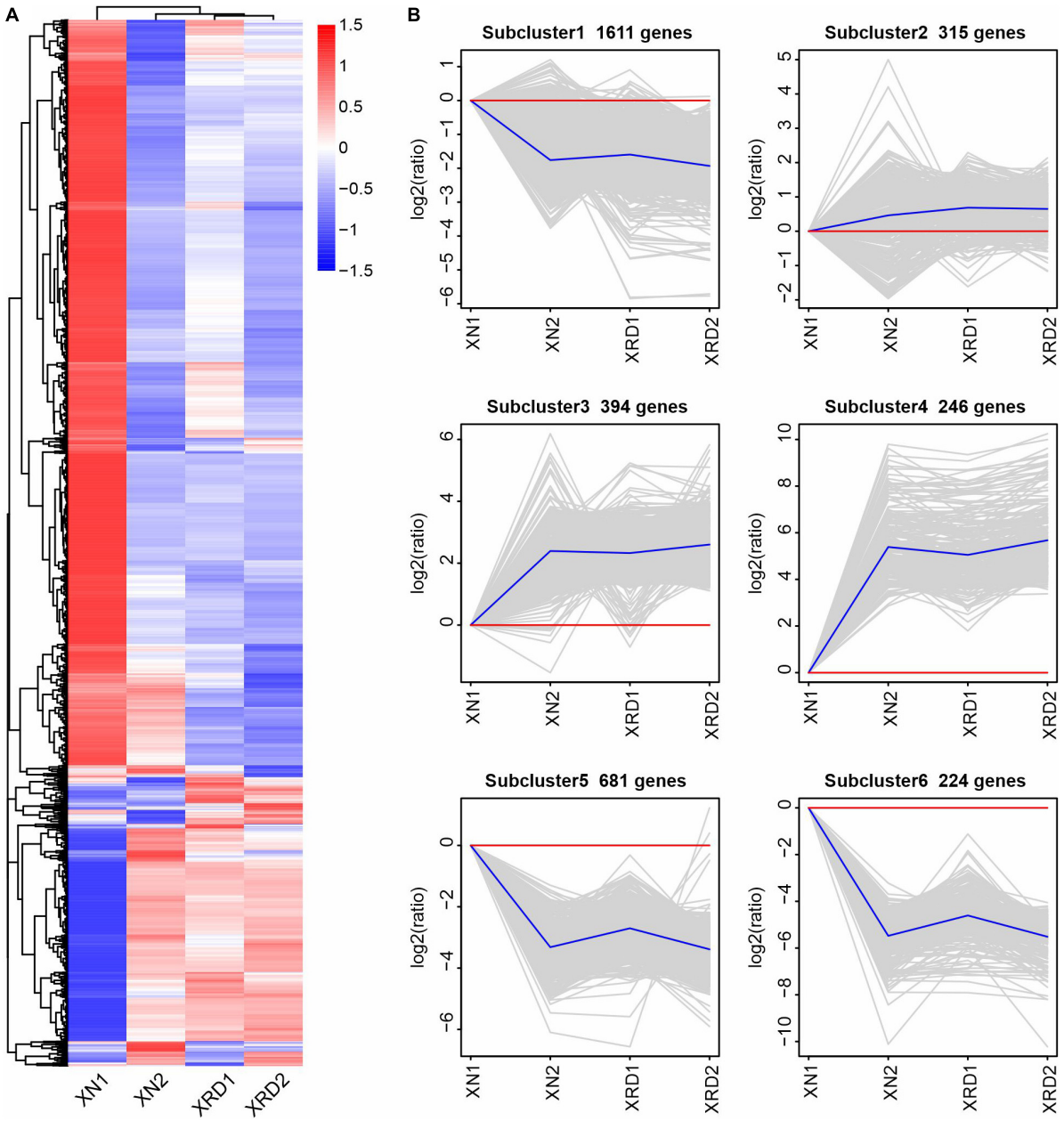


FIGURE 4
Differential gene H-cluster analysis and temporal dynamics analysis of the rapeseed transcriptome in XN and XRD at different time points. **(A)** Overall FPKM hierarchical cluster diagram normalized the $\log_{10}$(FPKM + 1) value and clusters it, red for high expression genes, and blue for low expression genes; The color is from red to blue, indicating that $\log_{10}$(FPKM + 1) is from large to small. **(B)** Log_2(ratio) diagram of temporal dynamics analysis. The gray lines in each subgraph represent the relative expression of a gene in a cluster under different experimental conditions. The blue line represents the average relative expression of all genes in this cluster under different experimental conditions. The *x*-axis represents the experimental conditions, and the y-axis represents the relative expression.

weight), and PN (pod number per plant) (Raboanatahiry et al., 2018). To further dissect the transcriptional profiles of genes involved in yield-related traits, we analyzed the expression profiles of genes related to SY (147), SW (189), and PN (28) during seed development in different areas (**Figure 5** and **Supplementary Table 9**). As shown in **Figure 5A**, the main significant DEGs related to SW included delta vacuolar processing enzyme (DELTA-VPE, BnaC05g31880D, BnaA05g18870D, BnaA01g25630D, BnaC04g29170D), HhH-GPD base excision DNA repair family protein (DME, BnaA10g25630D, BnaC09g50670D), ethylene-responsive transcription factor (WRI1, BnaA07g16350D, BnaA09g34250D,

BnaC08g25150D), and other DEGs, such as leucine-rich repeat transmembrane protein kinase (EXS/EMS1, BnaA10g23720D), amino acid permease 8 (AAP8, BnaC05g07760D), and MYB61 (BnaA08g26320D). The significant DEGs related to SY appeared mainly in hormone metabolism-related genes, such as the salicylic acid synthesis-degradation/C1-metabolism genes (BnaA01g15540D, BnaC01g18470D), alicyclic acid synthesis-degradation gene (BnaA03g10950D), gibberellin synthesis-degradation GA20 oxidase gene (BnaC02g01710D) and ethylene-related gene (RNA regulation of transcription Aux/IAA family, IAA7, BnaA03g36950D), and other significant DEGs, such as flavonoid chalcones naringenin-chalcone synthase gene (BnaA10g19670D) in secondary metabolism (**Figure 5B**). Among the PN-related genes shown in **Figure 5C**, there were three groups of significant DEGs, of which the first group contained delta (24)-sterol reductase genes (DWF1/DIM, BnaA05g19350D, BnaC05g32840D, BnaA06g02910D, BnaC01g33070D), the second group contained lysine-rich arabinogalactan protein (ATAGP19, BnaA07g27370D, BnaC06g30350D) and the third group includes 3-ketoacyl-CoA thiolase 5 (BnaC02g38800D, BnaA02g30470D) and 3-ketoacyl-CoA thiolase 2 (BnaC02g18580D, BnaC04g11470D).

Nine well-characterized genes, including *AUX1*, *CO*, *FT*, *FLC*, *PHYA*, *CRY2*, *AGL20*, *BRI1*, and *GAI*, were reported to be involved in key processes of crop yield determination, including the flowering process, light response, and plant hormone regulation (Geng et al., 2016, 2018; Lu et al., 2016, 2017; Pal et al., 2021). In this study, changes in these nine kinds of genes can be basically divided into three trends: the changes in the XN region are greater than the changes in the XRD region (standard changes); the changes in the XN region are less than the changes in the XRD region (special changes); and there is no change. Among them, PHYA (BnaC08g42660D) and CRY2 (BnaA10g02550D) are standard changes; AUX1 (BnaA05g06540D, auxin transport) and BRI1 (BnaA01g05490D, hormone metabolism, brassinosteroid signal transduction) also belong to standard changes, but GAI (BnaA02g12260D, hormone metabolism, gibberellin induced-regulated-responsive-activated) is unchanged. Among the other flowering-related genes, one FT (BnaA02g12130D) and AGL20 (BnaA03g56880D, RNA regulation of transcription MADS-box transcription factor family) belonged to standard changes, while CO (BnaA10g18430D) remained unchanged, but the remaining three genes, one FT (BnaC02g45250D) and two FLCs (BnaA02g00370D, BnaA03g02820D, RNA regulation of transcription MADS-box transcription factor family), belonged to special changes, which may play a special role in high yield in the XRD area (**Supplementary Table 9**).

To validate the expression profiles of yield-related genes from RNA-Seq data, 11 yield-related genes, which were differentially expressed at different time points in XN and XRD, were selected for qRT-PCR analysis, each with three bioreplicates (**Figure 6**). The qRT-PCR primers were designed based on rapeseed sequences from BRAD data, and their primer sequences are listed in **Supplementary Table 2**. We compared the results obtained from qRT-PCR with the results generated from RNA-Seq analysis of these yield-related genes, resulting in a correlation coefficient of R = 0.6 and P-value = 0.0031 (**Figure 6**), indicating that expression trends were consistent for all transcripts in both analyses.

Character seed filling stage by functional classification

To characterize the putative functions and pathways of the DEGs of each sample, we carried out the GO enrichment analysis of DEGs using GOseq. To acquire complete functional information, GO terms were assigned to each DEG. Among all the identified DEGs between XN2 vs. XN1, XRD2 vs. XRD1, XRD1 vs. XN1, and XRD2 vs. XN2, 1,425, 409, 1,311, and 705 were annotated with GO terms, respectively (**Figure 7** and **Supplementary Table 10**). Then, all DEGs were separated into upregulated and downregulated DEGs and functionally classified.

GO category analysis showed that DEGs identified in the XN area between the 30 vs. 40 DAF (XN2 vs. XN1) showed similar profiles for cellular component, biological process, and molecular function with XRD1 vs. XN1 (between XN and XRD area at the 30 DAF), with a high proportion of DEGs associated with nucleosome and chromosome in the cellular component. For the biological process category, DEGs were associated mainly with nucleosome assembly, chromatin assembly, and chromosome organization. For molecular function, protein heterodimerization activity and nutrient reservoir activity were the most highly represented categories (**Figures 7A,C**). In detail, among all the downregulated genes (1,349 out of 1,425 with GO annotations) between XN2 vs. XN1, 509, 167, 157, 130, 82, and 73 DEGs were assigned to heterocyclic compound binding (GO: 1901363), biosynthetic process (GO: 0009058), protein metabolic process (GO: 0019538), oxidation-reduction process (GO: 0055114), chromosome organization (GO: 0051276), and nucleosome assembly (GO: 0006334), respectively. In XRD1 vs. XN1, the downregulated genes (1,349 out of 1,425 with GO annotations) showed the same trend. In contrast, 464 upregulated genes between the 30 vs. 40 DAF in the XN area (XN2 vs. XN1) and 464 upregulated genes between the XN and XRD areas at the 30 DAF (XRD1 vs. XN1) showed different profiles. The 309, 80, 51, 72, 34, 22 up-DEGs related to biological process (GO: 0008150), oxidation-reduction process (GO: 0055114), small molecule metabolic process (GO: 0044281), oxidoreductase activity (GO: 0016491), nutrient reservoir activity (nutrient reservoir activity), lipid particle (GO: 0005811) in XN2 vs. XN1, while 296, 88, 75, 66, 47, 46, 33, 32, 27, 27, 22,19 up-DEGs related to

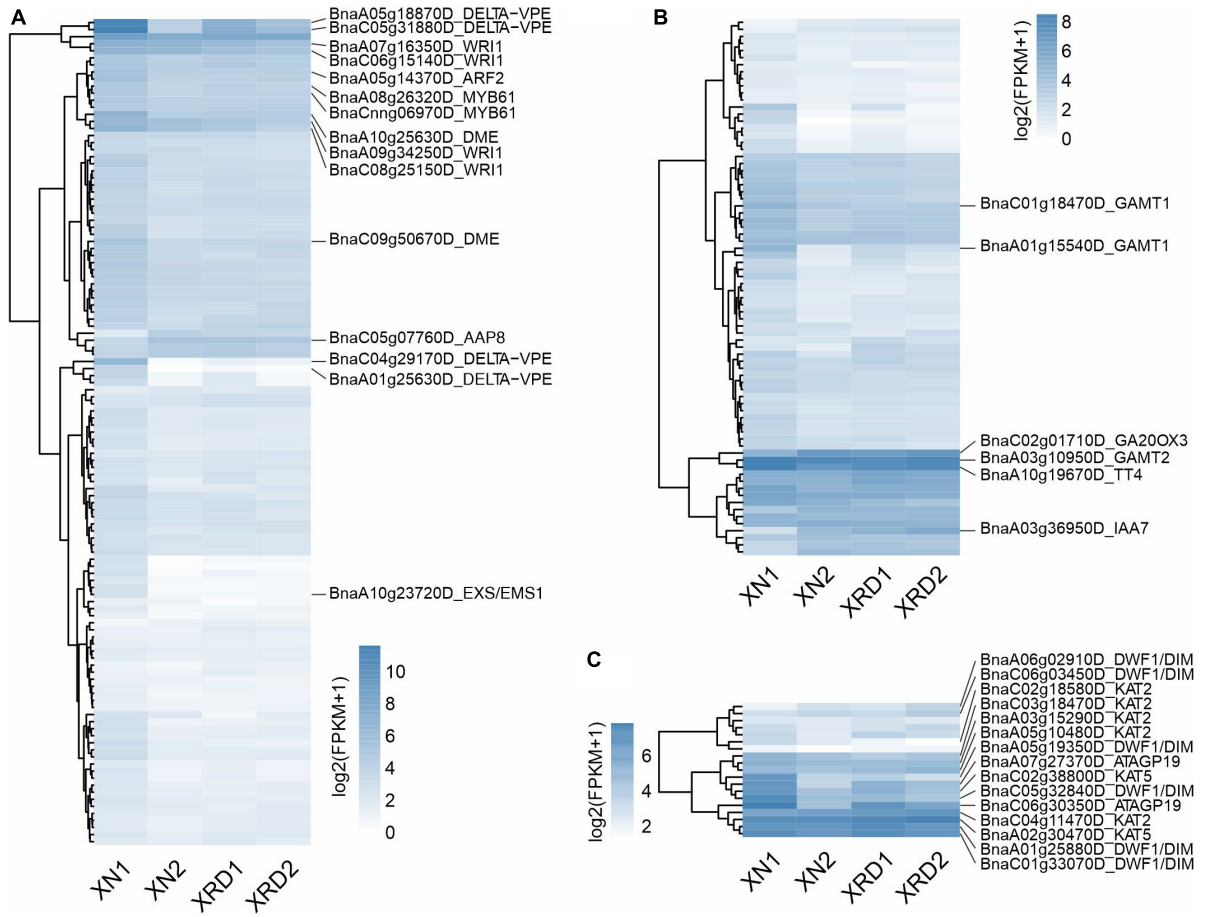


FIGURE 5
Expression profiles of genes implicated in yield traits. SW **(A)**, SY **(B)**, and PN **(C)**.

biological process (GO: 0008150), membrane (GO: 0016020), oxidation-reduction process (GO: 0055114), oxidoreductase activity (GO: 0016491), carbohydrate metabolic process (GO: 0005975), small molecule metabolic process (GO: 0044281), nutrient reservoir activity (GO: 0045735), lipid metabolic process (GO: 0006629), proteolysis (GO: 0006508), response to stimulus (GO: 0050896), lipid particle (GO: 0005811), cell wall organization or biogenesis (GO: 0071554) in XRD1 vs. XN1.

Most of the DEGs between the same XRD area (XRD2 vs. XRD1) were assigned to biological process, catalytic activity, hydrolase activity, and oxidoreductase activity, followed by cell wall, cell wall organization or biogenesis, reactive oxygen species (ROS) metabolic process, pectinesterase activity, carboxylic ester hydrolase activity, enzyme inhibitor activity, molecular function regulator, enzyme regulator activity, and endopeptidase activity (Figure 7B). In detail, among all the downregulated genes (50 out of 103 with GO annotations) in XRD2 vs XRD1, 18, 8, 8, 8, 6, 4, 4, and 4 DEGs were assigned to hydrolase activity (GO: 0016787), pectinesterase activity (GO: 0030599), carboxylic ester hydrolase activity (GO: 0052689), endopeptidase activity (GO: 0004175), enzyme inhibitor activity (GO: 0004857), photosystem (GO: 0009521), thylakoid (GO: 0009579), and photosynthesis (GO: 0015979), respectively. In contrast, 46 out of 53 upregulated genes in XRD2 vs. XRD1 were assigned to biological process (GO: 0008150), 14, 9, and 6 upregulated DEGs related to oxidoreductase activity (GO: 0016491), cellular component organization (GO: 0016043), and molecular function regulator (GO: 0098772), and 7, 4 upregulated DEGs related to cell wall organization (GO: 0071555) and cell wall (GO: 0005618), respectively.

Most of the DEGs between the same 40 DAF stage of XN and XRD seeds (XRD2 vs. XN2) were also assigned to biological process, metabolic process, macromolecular complex, cytoplasm, structural molecule activity followed by gene expression, peptide biosynthetic process, translation, ribosome, structural constituent of ribosome, and RNA binding (Figure 7D). In detail, among all the downregulated genes (179 out of 280 with GO annotations) in XRD2 vs XN2,

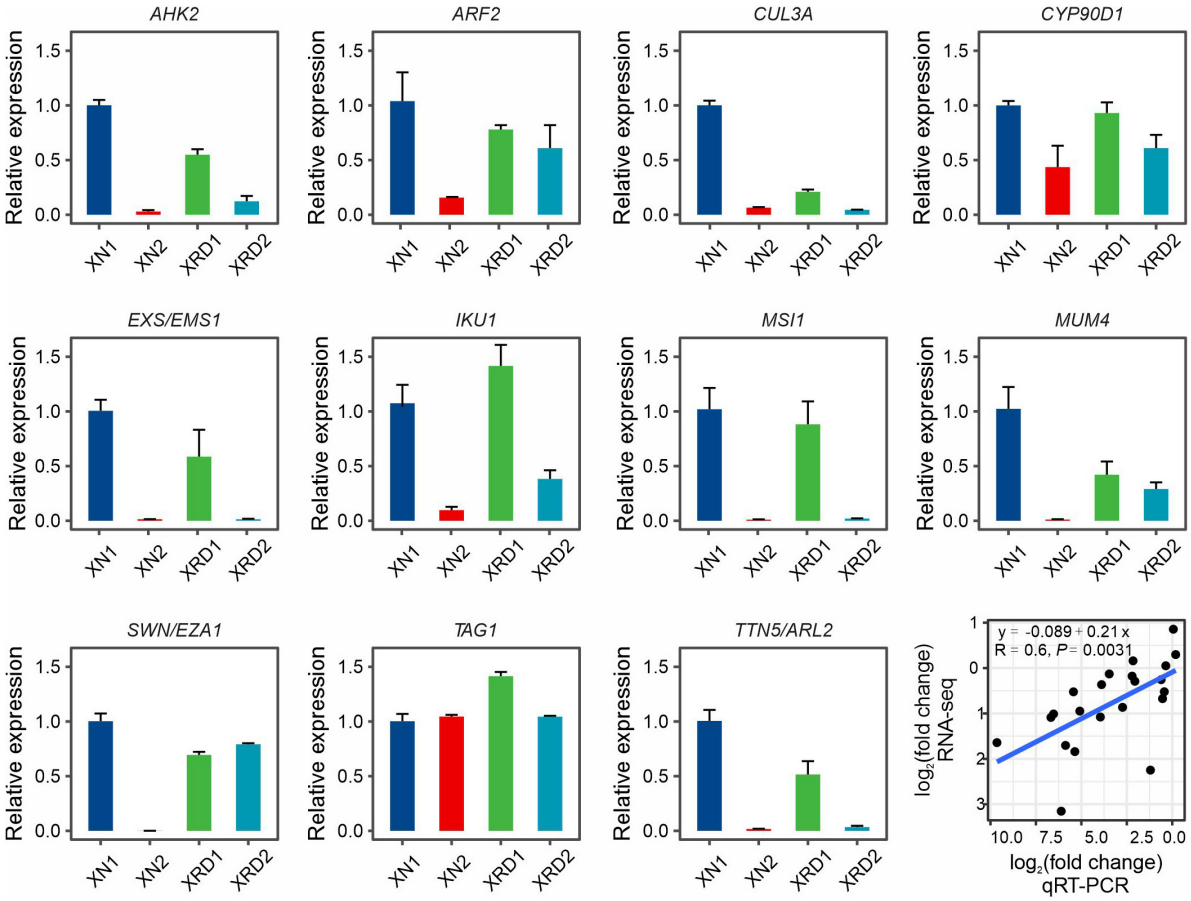


FIGURE 6
Expression profiles of yield-related genes in XN and XRD areas at 30 and 40 days based on qRT-PCR. Eleven yield-related genes with interests were selected. Correlation analysis showed that the qRT-PCR results were highly consistent with the RNA-seq results (P = 0.0031). Error bars indicate standard error (SE).

103, 78, 70, 61, 54, and 53 DEGs were assigned to the primary metabolic process (GO: 0044238), cellular biosynthetic process (GO: 0044249), gene expression (GO: 0010467), translation (GO: 0006412), structural molecule activity (GO: 0005198), and ribosome (GO: 0005840), respectively. In contrast, 64, 29, 22, 16, and 7 upregulated genes in XRD2 vs. XN2 were assigned to catalytic activity (GO: 0003824), hydrolase activity (GO: 0016787), membrane (GO: 0016020), proteolysis (GO: 0006508), and cell wall organization (GO: 0071555).

A pathway-based analysis will help us to further understand the biological functions of DEGs. To identify metabolic pathways in which DEGs were involved and enriched, the KEGG was used to analyze all the DEGs by identifying the top 20 most enriched KEGG pathways with KOBAS 2.0 (**Figure 8** and **Supplementary Table 11**). In the comparison of different times in the XN area (XN2 vs XN1), biosynthesis of secondary metabolites was the main enrichment pathway, followed by starch and sucrose metabolism, phenylpropanoid biosynthesis, amino sugar and nucleotide sugar metabolism, and photosynthesis. In detail, metabolic pathways, starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism, and photosynthesis were enriched mainly with downregulated DEGs in XN2 vs. XN1. In contrast, the upregulated DEGs in XN2 vs. XN1 were most significantly enriched in the biosynthesis of secondary metabolites, followed by phenylpropanoid biosynthesis and fatty acid metabolism (**Figure 8A**).

In the comparison of different times in the XRD area (XRD2 vs. XRD1), metabolic pathways were the main enrichment pathway, followed by photosynthesis-antenna proteins, cutin, suberin, and wax biosynthesis, photosynthesis, starch, and sucrose metabolism. In detail, metabolic pathways, photosynthesis-antenna proteins, photosynthesis, starch, and sucrose metabolism were mainly enriched with downregulated DEGs in XRD2 vs XRD1. In contrast, the upregulated DEGs in XRD2 vs. XRD1 were most significantly enriched in metabolic pathways, followed by biosynthesis of secondary

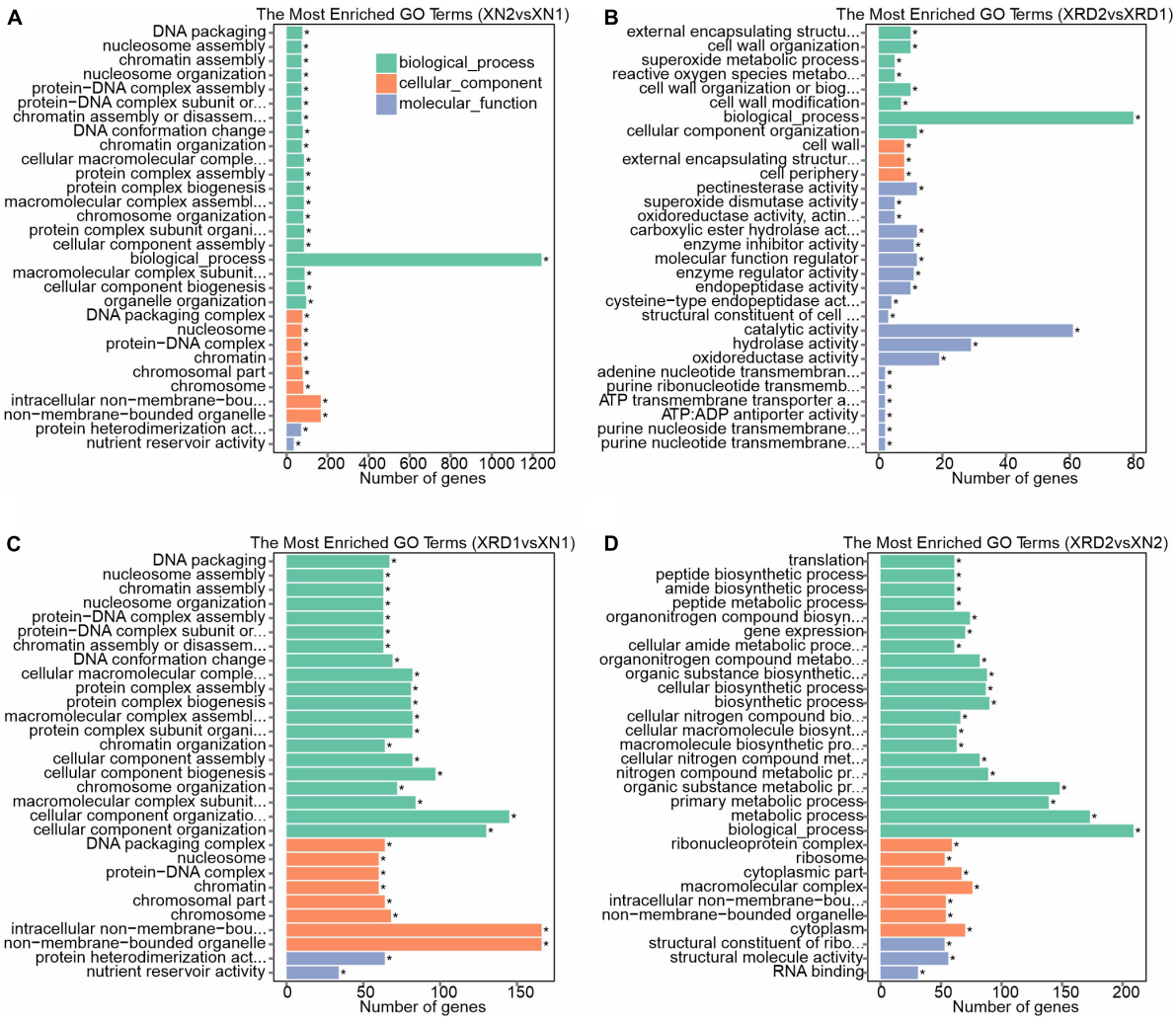


FIGURE 7
Gene function classification for DEGs identified in XN2 vs. XN1 **(A)**, XRD2 vs. XRD1 **(B)**, XRD1 vs. XN1 **(C)**, and XRD2 vs. XN2 **(D)**.

metabolites, cutin, suberine and wax biosynthesis, and peroxisome (Figure 8B).

In comparing different areas at 30DAF (XRD1 vs. XN1), starch and sucrose metabolism was the main enrichment pathway, followed by phenylpropanoid biosynthesis, ribosome biogenesis in eukaryotes, phagosome, arginine and proline metabolism, fatty acid metabolism. In detail, starch and sucrose metabolism, ribosome biogenesis in eukaryotes, phagosome, amino sugar and nucleotide sugar metabolism, and arginine and proline metabolism were enriched mainly with downregulated DEGs in XRD1 vs. XN1. In contrast, the upregulated DEGs in XRD1 vs. XN1 were most significantly enriched in the biosynthesis of secondary metabolites, followed by phenylpropanoid biosynthesis and tryptophan metabolism (Figure 8C).

In the comparison of different areas at 40DAF (XRD2 vs. XN2), the ribosome was the main enrichment pathway, followed by fatty acid metabolism, photosynthesis-antenna proteins, and circadian rhythm – plant. In detail, ribosome, fatty acid metabolism, biotin metabolism, circadian rhythm-plant, and RNA transport were noted to be more highly enriched with downregulated DEGs in XN2 vs. XN1. In contrast, the upregulated DEGs in XN2 vs. XN1 were most significantly enriched in metabolic pathways, followed by photosynthesis-antenna proteins, flavonoid biosynthesis, and alanine, aspartate, and glutamate metabolism (Figure 8D).

Rapid initiation into late seed development triggered by long sunshine is the key to high yield

A previous study has suggested that cellular activity during seed filling in *B. napus* begins with sugar mobilization,

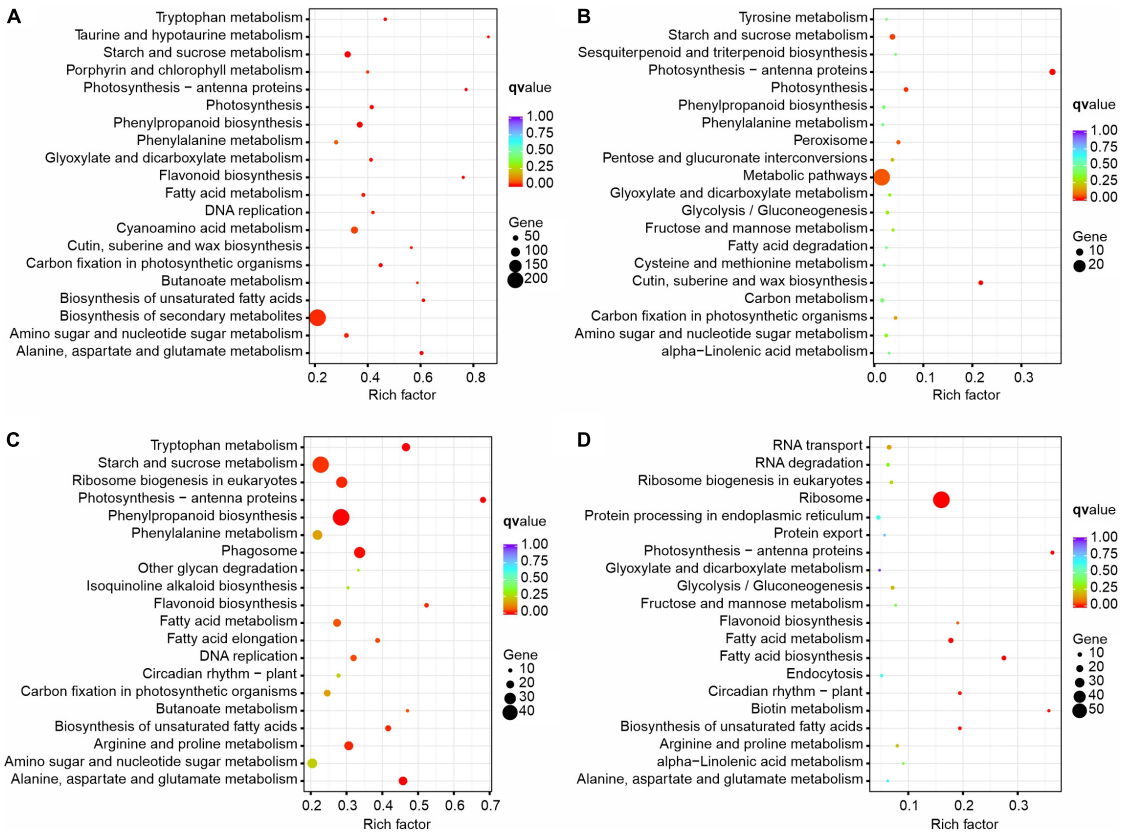


FIGURE 8
KEGG enrichment of DEGs identified from pairwise comparisons in XN2 vs. XN1 **(A)**, XRD2 vs. XRD1 **(B)**, XRD1 vs. XN1 **(C)**, and XRD2 vs. XN2 **(D)**.

followed by sequential surges in amino acid, lipid, and storage protein synthesis (Hajduch et al., 2006). They divided the expression trends into different functional subclasses based on three seed filling stages (**Figure 9A**). The first group included proteins expressed mainly at the early stages of seed filling, including proteins involved in glycolysis, respiration, metabolism of sugars, signal transduction, metabolism of amino acids, proteolysis, and defense. Proteins of the second group, involved in photosynthesis and lipid metabolism, exhibited the highest expression at the midpoint of seed filling. Finally, detoxification, seed maturation, and seed storage proteins were highly abundant at the end stage of seed filling.

Combining the above results from GO and KEGG pathway analyses and H-cluster analysis of DEGs, we can speculate that the seed developmental period in each sample is as follows (**Figure 9B** and 9C): 1) XN1 belongs to the early stage of expression, responsible mainly for starch and sucrose metabolism, nucleosome assembly, and chromatin assembly; 2) XN2 belongs to the midpoint stage of expression, accountable mainly for ribosome, fatty acid metabolism, photosynthesis, gene expression, and translation; 3) XRD1 belongs to the intermediate stage between the first and second stages of expression, responsible mainly for photosynthesis, starch and sucrose metabolism, cell wall organization or biogenesis, reactive oxygen species metabolic process, and enzyme regulator activity; and 4) XRD2 belongs to the end

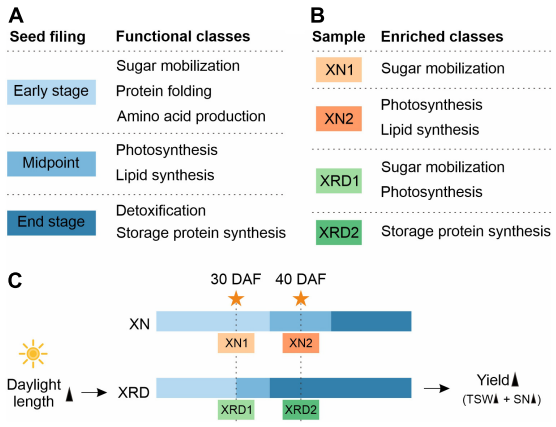


FIGURE 9
A proposed working model for rapeseed high yield in the unique plateau habitat.

stage of seed filling, in which DEGs were enriched in storage protein synthesis. Taken together, our data suggest that seed development in the XRD region occurred significantly earlier and lasted longer, which was caused by a relatively longer daylight length, resulting in a higher yield in the Xiangride area.

Discussion

Although numerous studies have focused on seed yield genetic improvement of rapeseed through quantitative trait locus (QTL) mapping, high-throughput sequencing, and genotyping techniques (Yang et al., 2012; Liu et al., 2015b; Zhao et al., 2016; Shah et al., 2018; Zheng et al., 2020; Khan et al., 2021; Hu et al., 2022), the comprehensive study of seed development and the mechanism contributing to high yield in the high-yield field, especially the Xiangride area, Tsaldam Basin of Qinghai Province in China, is poor, which is a constraint to our in-depth understanding of the environmental influences of this important trait and provides valuable information for future rapeseed breeding.

Relationship between high rapeseed yield and the unique eco-environment

Since Maskell and Mason proposed the source-sink theory to describe yield formation through the study of carbohydrate distribution in cotton plants in 1928, the methods of high yield from the source-sink point of view have often been discussed (Diepenbrock, 2000; Slafer, 2003). Subsequently, some crop traits were introduced into high yield formation; that is, crop yield is directly and multiply determined by yield component traits, including silique length (SL), seed weight/thousand seed weight (SW/TSW), SPS, and siliques per plant (SP), which are closely associated with seed yield improvement. Yield-related traits (such as biomass, harvest index, plant architecture, adaptation, and resistance to biotic and abiotic constraints) may also indirectly affect yield by affecting yield-component traits or by other unknown mechanisms (Diepenbrock, 2000; Chen et al., 2007). With the variety of crops, the yield components of high-yielding performance are different. In spring rapeseed, the high yield mainly showed that the number of pods per plant and 1,000-grain weight increased significantly, but the overall response was the synergistic effect of "source-intensive" and "sink-intensive" (Diepenbrock, 2000; Shi et al., 2009; Luo et al., 2015; Lu et al., 2016). In this study, we found that the SN and TSW in the XRD area were significantly greater than the SN and TSW in the XN area (**Figure 1**). This result showed that these two traits were the main factors resulting in high yield in the specific eco-environment of the XRD reproducing area.

The formation and potential development of high crop yields are greatly influenced by the distribution of photosynthetic assimilation among different organs and tissues. In the process of seed development, grain filling is an important stage to determine grain yield and is also a crucial and final stage of plant growth. Seed-filling duration is the morphology and physiological metabolism of rapeseed that will change dramatically in the process of grain filling, involving the supply of various components and precursors from the leaves into developing seeds, and diverse biochemical processes take place to synthesize carbohydrates, proteins, and lipids (Triboi et al., 2003; Barnabas et al., 2008; Awasthi et al., 2014). The yield and quality of grains are determined by the amount of organic matter synthesized or stored and the direction of transport, especially the photosynthate transport from the pod to seeds (Diepenbrock, 2000; Hajduch et al., 2006; Liu et al., 2015a). The production and transportation of photosynthetic assimilation are affected not only by the characteristics of the crops themselves but also by biotic or abiotic factors in the environment. Yield reflects the interaction of the environment with all growth and development processes that occur throughout the life cycle (Quarrie et al., 2006). In other words, yield is related to the environment; that is, seed development and even grain filling related to yield are also related to the environment. Many environmental conditions affect seed development, such as temperature, rainfall, light duration, and soil conditions (Barnabas et al., 2008; Huner et al., 2014; Fletcher et al., 2015; Macova et al., 2022). Some research results indicate that the high potential yield of irrigated rice in Yunnan is achieved mainly by intense incident solar radiation, which causes large biomass and N accumulation. The low nighttime temperature during the midgrowth stage was also suggested to be an important factor for considerable biomass accumulation and high grain yield in Yunnan (Ying et al., 1998a,b; Katsura et al., 2008). A high crop growth rate results from greater apparent canopy photosynthesis during daylight and less maintenance respiration at night (Sadras et al., 1993). Many experiments that compare the growth of a genotype in current and future projected elevated [CO_2] environments show that an increase in leaf photosynthesis is closely associated with similar increases in yield (Long et al., 2006). The physiological characteristics of high-yield crops in special ecological environment are reflected mainly in the improvement of net assimilation capacity, the extension of the functional period of photosynthetic organs after anthesis, the delay of photosynthetic organ senescence, and the large storage capacity, (Ying et al., 1998a,b; Long et al., 2006). In this study, the main advantage of the XRD area compared with the XN area is found to be that the sunshine length was longer, while other environmental factors (average temperature, rainfall, and average soil temperature of 20 cm) were lower than the environmental factors in the Xining area. Therefore, we suggest that the high yield of special habitats may be due to

the extension of sunshine time, and we can hypothesize that the specific eco-environment of the XRD area would lead to an increase in day length, which leads to the rise in the yield per plant.

Changes in yield candidate genes and metabolic pathways during seed development in *Brassica napus* L

Seed yield is a complex trait and can be improved by the direct component traits and the other indirect contributing traits. Both direct yield component traits and indirect yield-related traits are elaborately controlled by several genes and are highly influenced by environmental conditions (Van Camp, 2005; Shi et al., 2009; Ding et al., 2012; Zhao et al., 2016; Lu et al., 2017). The identification of yield candidate genes implies the detection of important genes for agricultural and economic quantitative traits (Raboanatahiry et al., 2018, 2022; Hu et al., 2021, 2022; Pal et al., 2021; Zheng et al., 2022). In the model plant *Arabidopsis thaliana*, 425 genes were collected from the TAIR website[2] to be related to flowering time, plant height, branch number, seed number, seed weight, and SY (Shi et al., 2009). In *B. napus*, several studies on yield candidate genes have previously been performed; for instance, Zhao et al. (2016) found four and two candidate genes for SY and SW traits (Shi et al., 2009; Zhao et al., 2016). Additionally, candidate genes for PN per plant (seven genes), branch PN (seven genes), and 1000 seed weights (one gene) were discovered (Lu et al., 2017). Moreover, a total of 1,562 genes in *B. napus* were homologous of 425 yield-related genes in *Arabidopsis thaliana,* and a total of 147 candidate genes were found inside regions of overlapping QTLs for nine traits, including PN, SW, and SY. Then, 147 SY-related, 189 SW-related, and 28 PN-related *Brassica napus* homologous genes of *Arabidopsis thaliana* yield-related genes were found (Raboanatahiry et al., 2018). Based on the construction of a quantitative genomic map of *Brassica napus*, Raboanatahiry further revealed 517 regions of overlapping QTLs that harbored 2,744 candidate genes and might affect multiple traits simultaneously (Raboanatahiry et al., 2022). Hu unraveled the genomic basis for the selection of adaptation and agronomic traits and identified 628 associated locus-related causative candidate genes for 56 agronomically important traits (including plant architecture and yield traits) through genome-wide association studies (Hu et al., 2022). In this study, we used these rapeseed candidate yield-related genes in our transcriptome data analysis and the results showed that the DEGs related to yield traits were generally scattered changes, which means that the high-yield environment may affect rapeseed yield indirectly.

2 http://www.arabidopsis.org/

Recent technical advancements in transcriptomics have provided new ways of addressing the temporal and spatial changes in gene expression associated with seed maturation and yield-determining traits (Harper et al., 2012; Liu et al., 2015a; Lu et al., 2016; Geng et al., 2018; Canales et al., 2021). Here, we carried out a transcriptome analysis of rapeseed seeds under XN and XRD high-altitude environmental conditions. Based on previous studies, the gene expression trends during seed filling in *B. napus* can be divided into three different functional stages, beginning with sugar mobilization, followed by sequential surges in photosynthesis and lipid metabolism, and finally, detoxification, seed maturation, and storage protein synthesis (Hajduch et al., 2006; Huang et al., 2013). In this study, compared with the filling development stage of rapeseeds, combined with transcriptome H-cluster analysis and GO and KEGG functional analyses, we found that the third development stage of rapeseeds in the XRD area, namely, the storage protein synthesis stage, was significantly ahead of schedule and lasted for a long time, resulting in a high XRD yield. Overall, based on the comparative analysis of environmental conditions and yields, we can infer that changes in environmental conditions, such as an increase in altitude, lead to an increase in sunshine length, and an increase in sunshine length leads to changes in the stage of rapeseed filling development and ultimately leads to an increase in rapeseed yield.

Conclusion

We conducted field yield analysis and transcriptome analysis in Xiangride, a special high-yield ecological area of rapeseed on the Qinghai Plateau. Compared with the yield and environmental factors in the Xining area, we found that the extension of daylight time in the Xiangride area may lead to an increase in rapeseed yield and significantly increases in seed weight (SW) and SN. Combined with transcriptome H-cluster analysis and GO and KEGG functional analyses, we can assume that the grain development of rapeseed in the Xiangride area is ahead of schedule and lasts for a long time, leading to the high-yield results in the Xiangride area, confirmed by the expression analysis by qRT-PCR of yield-related genes. These results suggest that the unique pattern of seed development in the XRD region, which developed ahead in the filling stage, was caused by a relatively longer daylight length, which is the key to high rapeseed yield in the Xiangride area. Our results provide valuable information for further exploring the molecular mechanism of high yield in special ecological environments and provide a useful reference for studying seed development characteristics in special regions of *B. napus*. At the same time, this study also provides a theoretical basis for how to improve high crop yield by adjusting the agricultural environment of rapeseed 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 below: Sequenced reads are available 573 in NCBI SRA (sequence read archive) under accession SRR 18827069 to SRR 18827072, 574 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA828284).

Author contributions

HX and RD designed the study. HX, HS, and RW performed experiments. HX, RD, and XJ analyzed the RNA-seq data and drafted the manuscript. HX, XJ, and RD revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant no. 31560052) and the Applied Basic Research Project of Qinghai Science and Technology Department of China (Grant no. 2015-ZJ-707 and 2019-ZJ-7049). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank the BGI Tech Solutions Co., Ltd. (BGI-Tech) for transcriptome sequencing.

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.2022.927418/full#supplementary-material

References

Awasthi, R., Kaushal, N., Vadez, V., Turner, N. C., Berger, J., Siddique, K. H. M., et al. (2014). Individual and combined effects of transient drought and heat stress on carbon assimilation and seed filling in chickpea. *Funct. Plant Biol.* 41, 1148–1167. doi: 10.1071/FP13340

Barnabas, B., Jager, K., and Feher, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 31, 11–38.

Basnet, R. K., Moreno-Pachon, N., Lin, K., Bucher, J., Visser, R. G., Maliepaard, C., et al. (2013). Genome-wide analysis of coordinated transcript abundance during seed development in different *Brassica rapa* morphotypes. *BMC Genomics* 14:840. doi: 10.1186/1471-2164-14-840

Borisjuk, L., Neuberger, T., Schwender, J., Heinzel, N., Sunderhaus, S., Fuchs, J., et al. (2013). Seed architecture shapes embryo metabolism in oilseed rape. *Plant Cell* 25, 1625–1640. doi: 10.1105/tpc.113.111740

Canales, J., Verdejo, J., Carrasco-Puga, G., Castillo, F. M., Arenas, M. A., and Calderini, D. F. (2021). Transcriptome analysis of seed weight plasticity in *Brassica napus*. *Int. J. Mol. Sci.* 22:4449. doi: 10.3390/ijms22094449

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

Chen, W., Zhang, Y., Liu, X., Chen, B., Tu, J., and Tingdong, F. (2007). Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F(2) populations. *Theor. Appl. Genet.* 115, 849–858. doi: 10.1007/s00122-007-0613-2

Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. *Field Crops Res.* 67, 35–49.

Ding, G., Zhao, Z., Liao, Y., Hu, Y., Shi, L., Long, Y., et al. (2012). Quantitative trait loci for seed yield and yield-related traits, and their responses to reduced phosphorus supply in *Brassica napus*. *Ann. Bot.* 109, 747–759. doi: 10.1093/aob/mcr323

Fletcher, R. S., Mullen, J. L., Heiliger, A., and Mckay, J. K. (2015). QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *J. Exp. Bot.* 66, 245–256. doi: 10.1093/jxb/eru423

Fu, Y., Wei, D., Dong, H., He, Y., Cui, Y., Mei, J., et al. (2015). Comparative quantitative trait loci for silique length and seed weight in *Brassica napus*. *Sci. Rep.* 5:14407. doi: 10.1038/srep14407

Geng, X., Dong, N., Wang, Y., Li, G., Wang, L., Guo, X., et al. (2018). RNA-seq transcriptome analysis of the immature seeds of two *Brassica napus* lines with extremely different thousand-seed weight to identify the candidate genes related to seed weight. *PLoS One* 13:e0191297. doi: 10.1371/journal.pone.0191297

Geng, X., Jiang, C., Yang, J., Wang, L., Wu, X., and Wei, W. (2016). Rapid identification of candidate genes for seed weight using the SLAF-Seq method in *Brassica napus*. *PLoS One* 11:e0147580. doi: 10.1371/journal.pone.0147580

Gupta, M., Bhaskar, P. B., Sriram, S., and Wang, P. H. (2017). Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. *Plant Cell Rep.* 36, 637–652. doi: 10.1007/s00299-016-2064-1

Hajduch, M., Casteel, J. E., Hurrelmeyer, K. E., Song, Z., Agrawal, G. K., and Thelen, J. J. (2006). Proteomic analysis of seed filling in Brassica napus.

Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. *Plant Physiol.* 141, 32–46. doi: 10.1104/pp.105.075390

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., et al. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802. doi: 10.1038/nbt.2302

Hay, J., and Schwender, J. (2011). Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (*Brassica napus* L.) embryos. *Plant J.* 67, 526–541. doi: 10.1111/j.1365-313X.2011.04613.x

Hu, D., Jing, J., Snowdon, R. J., Mason, A. S., Shen, J., Meng, J., et al. (2021). Exploring the gene pool of *Brassica napus* by genomics-based approaches. *Plant Biotechnol. J.* 19, 1693–1712. doi: 10.1111/pbi.13636

Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., et al. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. *Nat. Genet.* 54, 694–704. doi: 10.1038/s41588-022-01055-6

Hu, Y., Wu, G., Cao, Y., Wu, Y., Xiao, L., Li, X., et al. (2009). Breeding response of transcript profiling in developing seeds of *Brassica napus*. *BMC Mol. Biol.* 10:49. doi: 10.1186/1471-2199-10-49

Huang, D., Koh, C., Feurtado, J. A., Tsang, E. W., and Cutler, A. J. (2013). MicroRNAs and their putative targets in *Brassica napus* seed maturation. *BMC Genomics* 14:140. doi: 10.1186/1471-2164-14-140

Huner, N. P., Dahal, K., Kurepin, L. V., Savitch, L., Singh, J., Ivanov, A. G., et al. (2014). Potential for increased photosynthetic performance and crop productivity in response to climate change: role of CBFs and gibberellic acid. *Front. Chem.* 2:18. doi: 10.3389/fchem.2014.00018

Katsura, K., Maeda, S., Lubis, I., Horie, T., Cao, W., and Shiraiwa, T. (2008). The high yield of irrigated rice in Yunnan, China: 'A cross-location analysis'. *Field Crops Res.* 107, 1–11.

Khan, S. U., Saeed, S., Khan, M. H. U., Fan, C., Ahmar, S., Arriagada, O., et al. (2021). Advances and challenges for QTL analysis and GWAS in the plant-breeding of high-yielding: a focus on rapeseed. *Biomolecules* 11:1516. doi: 10.3390/biom11101516

Leon, J. (1993). The importance of crop physiology for the breeding of oilseed rape. *Fett* 95, 283–287.

Li, N., Shi, J., Wang, X., Liu, G., and Wang, H. (2014). A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (*Brassica napus* L.). *BMC Plant Biol.* 14:114. doi: 10.1186/1471-2229-14-114

Liu, H., Yang, Q., Fan, C., Zhao, X., Wang, X., and Zhou, Y. (2015a). Transcriptomic basis of functional difference and coordination between seeds and the silique wall of *Brassica napus* during the seed-filling stage. *Plant Sci.* 233, 186–199. doi: 10.1016/j.plantsci.2015.01.015

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

Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330.

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., et al. (2017). Genome-wide association and transcriptome analyses reveal candidate genes underlying yield-determining traits in *Brassica napus*. *Front. Plant Sci.* 8:206. doi: 10.3389/fpls.2017.00206

Lu, K., Xiao, Z., Jian, H., Peng, L., Qu, C., Fu, M., et al. (2016). A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in *Brassica napus*. *Sci. Rep.* 6:36452. doi: 10.1038/srep36452

Luo, X., Ma, C., Yue, Y., Hu, K., Li, Y., Duan, Z., et al. (2015). Unravelling the complex trait of harvest index in rapeseed (*Brassica napus* L.) with association mapping. *BMC Genomics* 16:379. doi: 10.1186/s12864-015-1607-0

Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., et al. (2017). Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585.

Macova, K., Prabhullachandran, U., Stefkova, M., Spyroglou, I., Pencik, A., Endlova, L., et al. (2022). Long-term high-temperature stress impacts on embryo and seed development in *Brassica napus*. *Front. Plant Sci.* 13:844292. doi: 10.3389/fpls.2022.844292

Mason, A. S., and Snowdon, R. J. (2016). Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. *Plant Biol.* 18, 883–892. doi: 10.1111/plb.12462

Pal, L., Sandhu, S. K., Bhatia, D., and Sethi, S. (2021). Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus* subsp. napus). *Physiol. Mol. Biol. Plants* 27, 1933–1951. doi: 10.1007/s12298-021-01060-9

Quarrie, S., Pekic Quarrie, S., Radosevic, R., Rancic, D., Kaminska, A., Barnes, J. D., et al. (2006). Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J. Exp. Bot.* 57, 2627–2637. doi: 10.1093/jxb/erl026

Quijada, P. A., Udall, J. A., Lambert, B., and Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theor. Appl. Genet.* 113, 549–561. doi: 10.1007/s00122-006-0323-1

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.* 9:1127. doi: 10.3389/fpls.2018.01127

Raboanatahiry, N., Chao, H., He, J., Li, H., Yin, Y., and Li, M. (2022). Construction of a quantitative genomic map, identification and expression analysis of candidate genes for agronomic and disease-related traits in *Brassica napus*. *Front. Plant Sci.* 13:862363. doi: 10.3389/fpls.2022.862363

Sadras, V. O., Hall, A. J., and Connor, D. J. (1993). Light-associated nitrogen distribution profile in flowering canopies of sunflower (*Helianthus annuus* L.) altered during grain growth. *Oecologia* 95, 488–494. doi: 10.1007/BF00317432

Shah, S., Karunarathna, N. L., Jung, C., and Emrani, N. (2018). An APETALA1 ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biol.* 18:380. doi: 10.1186/s12870-018-1606-9

Shi, J., Li, R., Qiu, D., Jiang, C., Long, Y., Morgan, C., et al. (2009). Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics* 182, 851–861. doi: 10.1534/genetics.109.101642

Shi, S. B., Chen, W. J., Shi, R., Li, M., Zhang, H. G., and Sun, Y. N. (2014). PS II photochemical efficiency in flag leaf of wheat varieties and its adaptation to strong sun- light intensity on farmland of Xiangride in Qinghai Province, Northwest China. *Ying Yong Sheng Tai Xue Bao* 25, 2613–2622.

Slafer, G. A. (2003). Genetic basis of yield as viewed from a crop physiologist's perspective. *Ann. Appl. Biol.* 142, 117–128.

Su, T., and Pan, Z. (1981). An analysis of the physiological features of the higher yielding ability of spring wheat in the Xiangride farm, Qinghai province. *Acta Agron. Sin.* 1, 19–26.

Triboi, E., Martre, P., and Triboi-Blondel, A. M. (2003). Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. *J. Exp. Bot.* 54, 1731–1742. doi: 10.1093/jxb/erg183

Udall, J. A., Quijada, P. A., Lambert, B., and Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor. Appl. Genet.* 113, 597–609. doi: 10.1007/s00122-006-0324-0

Van Camp, W. (2005). Yield enhancement genes: seeds for growth. *Curr. Opin. Biotechnol.* 16, 147–153.

Yang, P., Shu, C., Chen, L., Xu, J., Wu, 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

Ying, J., Peng, S., He, Q., Yang, H., Yang, C., Visperas, R. M., et al. (1998a). Comparison of high-yield rice in tropical and subtropical environments: I. Determinants of grain and dry matter yields. *Field Crops Res.* 57, 71–84.

Ying, J., Peng, S., Yang, G., Zhou, N., Visperas, R. M., and Cassman, K. G. (1998b). Comparison of high-yield rice in tropical and subtropical environments: II. Nitrogen accumulation and utilization efficiency. *Field Crops Res.* 57, 85–93.

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

Zhao, X., Yu, K., Pang, C., Wu, X., Shi, R., Sun, C., et al. (2021). QTL analysis of five silique-related traits in *Brassica napus* L. Across multiple environments. *Front. Plant Sci.* 12:766271. doi: 10.3389/fpls.2021.766271

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., et al. (2017). Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 8:1246. doi: 10.3389/fpls.2017.01246

Zheng, M., Terzaghi, W., Wang, H., and Hua, W. (2022). Integrated strategies for increasing rapeseed yield. *Trends Plant Sci.* S1360-1385, 00080–00082.

Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., et al. (2020). Knockout of two BnaMAX1 homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnol. J.* 18, 644–654. doi: 10.1111/pbi.13228

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fpls.2022.927662

OPEN ACCESS

EDITED BY
Naser A. Anjum,
Aligarh Muslim University, India

REVIEWED BY
Helene S. Robert Boisivon,
Central European Institute of Technology (CEITEC), Czechia
Benbo Xu,
Yangtze University, China

*CORRESPONDENCE
Shuijin Hua
sjhua1@163.com

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

RECEIVED 24 April 2022
ACCEPTED 17 August 2022
PUBLISHED 09 September 2022

CITATION
Hao P, Lin B, Ren Y, Hu H, Xue B, Huang L and Hua S (2022) Auxin-regulated timing of transition from vegetative to reproductive growth in rapeseed (*Brassica napus* L.) under different nitrogen application rates. *Front. Plant Sci.* 13:927662. doi: 10.3389/fpls.2022.927662

Auxin-regulated timing of transition from vegetative to reproductive growth in rapeseed (*Brassica napus* L.) under different nitrogen application rates

Pengfei Hao[1], Baogang Lin[1], Yun Ren[2], Hao Hu[3], Bowen Xue[1], Lan Huang[1] and Shuijin Hua[1]*

[1]Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]Huzhou Agricultural Science and Technology Development Center, Huzhou, China, [3]Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

Accelerating the differentiation of floral meristem (FM) from shoot apical meristems (SAM) which determines the conversion from vegetative to reproductive growth is of great significance for the production of rapeseed (*Brassica napus* L.). In this research, the mechanisms of different nitrogen (N) application rates (low N, N1; normal N, N2; and high N, N3) on different FM development stages triggering the regulation of FM differentiation genes through the auxin biosynthetic and signal transduction were investigated. We found that the stage of FM differentiation, which was identified through a stereomicroscope and scanning electron microscope, came 4 and 7 days earlier under high N rate than under normal and low N levels, with the seed yield increased by 11.1 and 22.6%, respectively. Analysis of the auxin and its derivatives contents showed that the main biosynthesis way of auxin was the indole acetaldehyde oxime (IAOx) pathway, with 3-Indole acetonitrile dramatically accumulated during FM differentiation. At the same time, an obvious decrease of IAA contents at each FM differentiation stage was detected, and then gradually rose. Results of the expression of genes involved in auxin biosynthesis, auxin signaling transduction, and FM identification under five FM differentiation stages and three nitrogen application rates showed that genes involved in auxin biosynthesis were regulated before the FM differentiation stage, while the regulation of FM identity genes appeared mainly at the middle and later periods of the five stages, and the regulation level of genes varied under different N rates. Taken together, a high nitrogen rate could accelerate the initiation of FM differentiation, and auxin involved a lot in this regulation.

KEYWORDS

auxin biosynthesis, floral meristem, nitrogen application, shoot apical meristem, yield

Introduction

Appropriate timing of the transition from vegetative to reproductive growth is highly correlated with crop flowering time and yield (Narnoliya et al., 2019). The morphologic landmark of this change is a protrusion in the peripheral zone, which is called flower primordium (Carles and Fletcher, 2003). The differentiation of flower primordia derives from a population of cells called shoot apical meristem (SAM) (Tooke and Battey, 2003). The structure of the SAM had been elucidated in many plant species, which includes the central zone containing the stem cells, the peripheral zone initiating primordia, and the rib zone producing the internal part of the stem (Vernoux et al., 2010).

Extensive genetic studies identified many important genes controlling the transition from vegetative to reproductive growth. One of the regulating pathways of SAM differentiation is a system centered on a transcription factor WUSCHEL (WUS) (Hamada et al., 2000). The *WUS* signaling pathway was negatively regulated by *CLAVATA3* (*CLV3*), which encodes a small secreted signaling peptide (Xin et al., 2017). During SAM differentiation, the opposite roles between *CLV3* and *SHOOTMERISTEMLESS* (*STM*) were found (Nidhi et al., 2021). A combination of *STM* and *CUP-SHAPED COTYLEDON* (*CUC*)*1* family prevents new organ initiation (Balkunde et al., 2017). On the other hand, some key genes such as *LEAFY* (*LFY*) promote the outgrowth of flower primordium with the interaction of *WUS* or *AGAMOUS* (*AG*) (Traas, 2018). Therefore, the equilibrium between these two opposite gene systems is the main endogenous force for SAM differentiation. Other regulation systems for SAM differentiation were also identified. For example, post-transcriptional regulation via microRNAs (miRNAs) such as miR165/166 had important roles in SAM differentiation (Zhou et al., 2015). The most difficulties of these regulation systems are how the regulatory network is coordinated.

Among the mechanisms of feedback loops to sustain and restrict stem cell activities, one clue showed that auxin had an essential role in regulating the formation of flower primordium (Luo et al., 2018; Shi and Vernoux, 2022). The explanation for auxin regulation on SAM differentiation is mainly due to the unbalance of auxin concentration in different zones, which was affected by different auxin transporter systems such as PIN protein and AUX/LAX influx carrier (Abley et al., 2016). Some models such as the flux-based and up-the-gradient models have been advised to explain the transportation and allocation of auxin in different cell zones (Bilsborough et al., 2011). However, the elaborate modulation system on SAM differentiation by auxin is still on the way.

In addition to genetic controlling SAM differentiation, environmental and agronomic practices such as low temperature and nitrogen application also have important roles in the differentiation (Olas et al., 2019). Although nitrogen application can affect crop growth and development, there are few evidence of the relationship between nitrogen application and SAM differentiation (Olas et al., 2019). It is well known that there is a close relationship between nitrogen application rate and auxin biosynthesis through the tryptophan pathway (Zhao, 2011; Fu et al., 2020; Gu et al., 2022). And auxin has important regulatory role in SAM differentiation. Therefore, our hypothesis in this research is that nitrogen application rate affects SAM differentiation through auxin mediation. In the present study, we first analyzed the effect of nitrogen application rate on the timing of SAM differentiation in rapeseed. And then, we dissected the auxin concentration in the SAM with various differential stages. Finally, we also assayed the gene expression of auxin biosynthesis, transportation, and SAM differentiation pathway through transcriptome. The research will enrich our scientific knowledge on the molecular regulation of SAM differentiation by nitrogen application rate through auxin regulation.

Materials and methods

Plant materials and experimental design

The field trials were conducted during the growing seasons of 2020–2021 and 2021–2022 at the Zhejiang Academy of Agricultural Sciences, Hangzhou, China. Commercial rapeseed (*Brassica napus* L.) variety zheyou 50 was used as plant material. The soil type in the experimental station is loamy clay (loamy, mixed, and thermic Aeric Endoaquepts). The previous crop was rice. Before sowing, fertilizers including calcium superphosphate, potassium oxide, and borax were manually applied at the rate of 375, 120, and 15 kg ha^{-1}, respectively as a basal fertilizer dose. Approximately five to six rapeseed seeds were directly sown into the soil in each shallow hole at a depth of approximately 3 cm in each plot. After sowing, the plot was covered by a layer of soil and then sprayed 750 mL of s-metolachlor (96%, v/v) with 1,500 kg of water per hectare to control the weed. The excessive seedlings were removed and thinned to one plant in each hole after 1 month. The field was not irrigated during rapeseed growing seasons. The growth condition, mainly including mean temperature and rainfall, was illustrated in Supplementary Figure S1.

Experimental design

Nitrogen fertilizer (urea) was applied before sowing with three levels, which were low nitrogen level (N1, 120 kg ha^{-1}), optimal nitrogen level (N2, 240 kg ha^{-1}), and high nitrogen level (N3, 360 kg ha^{-1}). Under each treatment, 70% of the urea was applied into the soil before sowing and 30% of the urea was applied on top during the budding stage. A completely

randomized block method was used with three replicates and the area of each block was 18 m^2. The density of plants in each treatment was 120,000 plants ha^{-1} with the same management during the whole growing season.

Identification and determination of floral meristem (FM) initiation time and morphology and sampling

About 30 days before the initiation of floral meristem differentiation, all rapeseed plants at a random area of 2 m^2 in each plot were sampled (about 25 plants per plot) and the morphology of SAM was observed after peeling off all the visible blades under a stereoscopic microscope (VHX-950F, Japan) every 5 d (Zhang et al., 2016). The appearance of ball-shaped flower meristems (FM) in inflorescence meristems (IM) is determined as the initiation of FM differentiation of each plant, and 60% of plants entering initiation were determined as FM differentiation (FMd) time (Figures 1B,C).

The morphology of SAM at 5 days before N3 came into FMd (T0, November 25th), N3 just came into FMd (T1, November 30th), N2 entered into FMd (T2, December 4th), and N1 entered into FMd (T3, December 7th) and 5 d after N1 entered into FMd (T4, December 12th) were observed under scanning electron microscopy according to Vernoux et al. (2000), and the samples of SAM at these five stages were collected and kept at −80°C for RNA-seq and plant hormone determination.

RNA extraction, library construction, sequencing, and analysis

Samples of SAM at T0, T1, T2, T3, and T4 stages under three nitrogen application levels were used as materials and total RNA was extracted using the polysaccharide and polyphenol total RNA isolation kit (Bioteke, Beijing, China). The libraries were constructed according to the manufacturer's instructions and performed on Illumina Nova-seq 6,000 system for sequencing, and HISAT2 was used to align all clean reads against the reference genome *Brassica napus* (Fu et al., 2018). The expression levels of genes were determined according to FPKM values. $|\log_2 \text{FoldChange}|>1$ and P-value ≤ 0.05 were determined as differentially expressed genes (DEGs) using the DEseq2 methods and edgeR program (Liu et al., 2020). Genes related to auxin biosynthesis, auxin signaling transduction, and floral meristem differentiation were focused. The expression of each gene at 5 stages × 3 nitrogen levels were demonstrated as a heatmap after data centralization and standardization by R programming language.

Determination of auxins contents by ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS)

Auxin contents of SAM at T0, T1, T2, T3, and T4 stages under three nitrogen application levels were determined through UPLC-MS/MS methods. The pretreatment of samples was performed according to Li et al. (2016). Supernatants were subjected to UPLC-MS/MS system, and then Multiple Reaction Monitoring (MRM), Analyst v.1.6.3, and MultiQuant v.3.0.3 were utilized to quantitate and quantify the raw data (Šimura et al., 2018). The contents of each auxin at 5 stages × 3 nitrogen levels were shown as a heatmap after data centralization and standardization by R programming language.

Statistics

Data analysis was performed using IBM SPSS v.22.0 statistical software. Duncan's multiple range test (DMRT) was used to evaluate significant treatment effects at the significance level of $P \leq 0.05$.

Results

Higher nitrogen application rate promoted the differentiation of FM and yield

The number of effective siliques and yield were significantly improved by the nitrogen application rate (Figure 1A). The number of effective siliques was 198 per plant on average under N3, which was 11.1 and 22.6% higher than under N2 and N1, respectively. The yield was also increased by 26.1 and 50.5% as compared to N2 and N1.

Correlation analysis showed that the ratio of FM differentiation was highly correlated with seed yield, with the correlation coefficient reaching 0.978, 0.992, and 0.977 at three sampling stages, respectively (Supplementary Figure S2). Given the close correlation between the timing of FM differentiation (FMd) and yield (effective siliques), we analyzed the timing of FM differentiation under different nitrogen application rates. Under optimal nitrogen rate (N2) treatment, FM differentiation of rapeseed initiated on December 4th (T2), 3 days earlier than under low nitrogen rate (N1), which initiated on December 7th (T3), and 4 days later than under high nitrogen rate (N3), which initiated on November 30th (T1) (Figure 1C).

In addition to the differences in the timing of the transition from vegetative to reproductive, plant height, leaf chlorophyll content, and leaf numbers were also analyzed when plants entered FMd under N3 treatment. As the nitrogen application

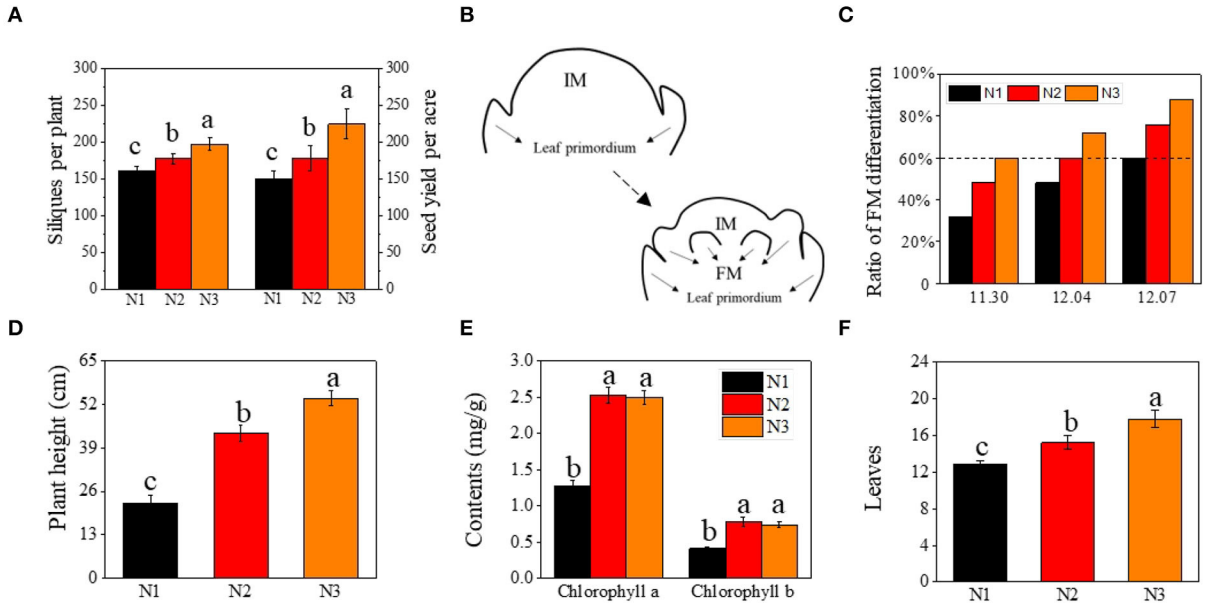


FIGURE 1
Effects of low N (N1), normal N (N2), and high N (N3) rates on silique number, and yield of rapeseed, floral meristems differentiation, plant height, chlorophyll contents, and leaves. **(A)** silique number and seed yield per acre; **(B)** the morphogenesis of floral meristems initiation; **(C)** ratio of floral meristems differentiation (%) under different N rates at different sampling stage; **(D–F)** represent plant height (cm), chlorophyll contents (mgg^{-1}), and leaves, respectively; FM represented floral meristem. IM represented inflorescence meristem. Error bars represented SD value. Different lower cases represented significance at $p \leq 0.05$.

rate increased, plant height under N3 was 24.2% and 139% higher than that under N2 and N1, respectively (Figure 1D). As for leaf chlorophyll a and chlorophyll b content, there were no significant differences between N2 and N3 treatments, however, both nitrogen application rates were significantly higher than that under N1 treatment, which increased by 98.8 and 90.2% under N2 treatment, and 96.3 and 79.8% at N3, respectively, as compared to N1 treatment (Figure 1E). For the number of leaves, it increased by 16.9 and 38.0% under N3 treatment when compared with N2 and N1, respectively (Figure 1F).

Morphology of FMd under different nitrogen application rates

For the morphological changes during FMd, the process of FMd was first imagined by stereomicroscope. One set of the process was illustrated in Figure 2A. The process can be roughly divided into five stages. In brief, in stage 1, plants were before floral meristem differentiation with leaf primordia (LP) around the SAM (Figure 2A, 1∼2). Stage 2, SAM developed into inflorescence meristems (IM) quickly, and plants were entering floral meristem differentiation with one or two protrusion(s) on the IM (Figure 2A, 3). Stage 3, floral primordium burst around the IM (Figure 2A, 4∼7). Stage 4, floral primordial rapid accumulation corresponded with the floral organ development (Figure 2A, 8). Stage 5, the main inflorescence was established (Figure 2A, 9∼10).

To further understand the morphological differences in the timing of the transition from vegetative to reproductive growth caused by different nitrogen application rates, a scanning electron microscope was used to have a clear presentation. At the T0 stage, the SAM was not differentiated with only leaf primordia under three nitrogen application rates (Figure 2B i, vi, and xi). At the T1 stage, SAM developed into IM with several bulbs surrounded under N3 treatment, suggesting the initiation of FM differentiation. However, there were no obvious protrusions under N1 and N2 treatments (Figure 2B ii, vii, and xii). At the T2 stage, there were about eight bulbs observed around the IM under N3 and two protrusions under N2 treatment, while, no protrusion was found around the IM under N1 treatment at this stage (Figure 2B iii, viii, and xiii). At the T3 stage, many bulbs were distributed around the SAM and some bulbs started to develop into flowers under N3 treatment, while considerable bulbs were around the IM but no differentiated bulbs under N2 treatment. There were two bulbs observed around the IM under N1 treatment indicating the start of FM differentiation at this stage (Figure 2B iv, ix, and xiv). At the T4 stage, rapeseed entered FM differentiation under all nitrogen application rates only with the different differentiation rates (Figure 2B v, x, and xv). The morphologic result clearly

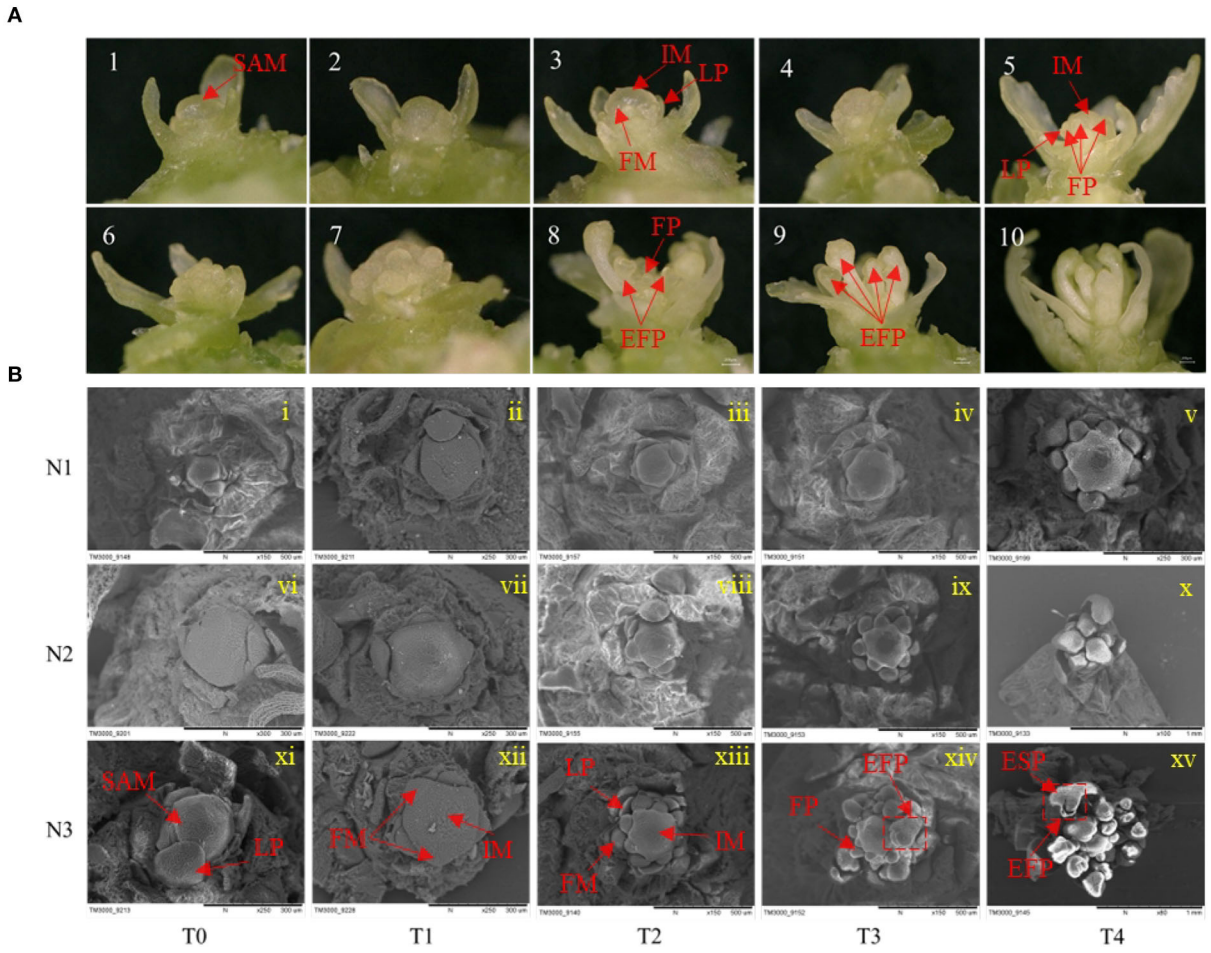


FIGURE 2
Effects of low N (N1), formal N (N2), and high N (N3) rates on the morphology of shoot apical meristems and floral meristems. **(A)** display of 10 consecutively different images of the development of rapeseed inflorescence and floral meristems under a stereoscope. **(B)** Scanning electron microscope images of inflorescence and floral meristems at T0, T1, T2, T3, and T4 sampling stages under low N (N1), normal N (N2), and high N (N3) rates. EFP, elongated floral primordium; ESP, elongated sepal primordium; FM, floral meristem; FP, floral primordium; IM, inflorescence meristem; LP, leaf primordium; SAM, shoot apical meristem.

showed that a higher nitrogen application rate accelerated the timing of the transition from vegetative to reproductive growth in rapeseed.

Endogenous auxin contents in rapeseed SAM were influenced by different nitrogen application rates

To uncover the reason for earlier timing of the transition from vegetative to reproductive growth promoted by higher nitrogen application rate, the variation tendency of endogenous auxin contents of SAM at five stages under three nitrogen rates were assayed. By utilizing UPLC-MS/MS method, four auxin metabolites involved in auxin biosynthesis were detected, which were tryptophan (Trp), 3-Indole acetonitrile (IAN), 3-Indole acetamide (IAM), and Indole-3-acetic acid (IAA) involving in two auxins biosynthetic pathways, respectively, which were indole acetaldehyde oxime pathway (IAOx pathway) and indole acetamide pathway (IAM pathway). Among those products, Trp and IAN contents were dominant as compared to other ones. The contents of Trp were higher before each FM initiation stage under different nitrogen levels, then decreased sharply at the FMd stage and gradually increased whereafter. The Trp content at N3-T1 was significantly lower than that at N1-T3 (both at the FM initiation stage), which showed a 24.1% decrement of content. However, no significant difference was found between N2 and N3 or N1 and N2, at each FM initiation stage. The contents of IAN were dramatically higher than other IAA derivatives and showed a trend of decline with fluctuation, while the contents of IAN at T4 were largely decreased than other stages. Aimed at the FM initiation stage, IAN content at N1-T3 was 234.8 and 360.6% significantly higher than those at N2-T2

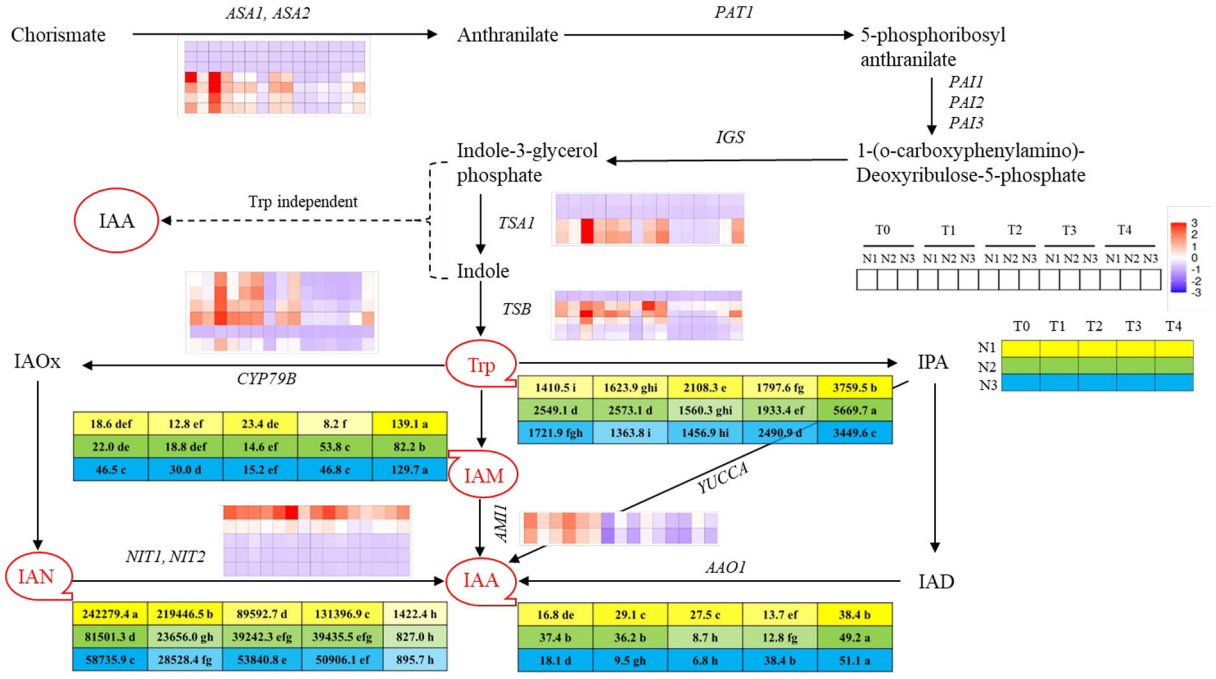


FIGURE 3
Overview of auxin and derivatives and genes involved in auxin biosynthesis pathway at T0, T1, T2, T3, and T4 stages under low N (N1), normal N (N2), and high N (N3) rates. IAA, Indole-3-acetic acid; IAN, 3-Indole acetonitrile; IAM, 3-Indole acetamide; Trp, tryptophan. Red circles with tails represented different auxin-derivatitves and the 5 × 3 matrix represented the contents under 3 N rates at five sampling stages. Color bars represent the trends of expression levels after centralization and standardization by the R programming language, red color represents higher expression, while purple represents lower expression relatively. Tables near Trp, IAM, IAN, and IAA represent each content (ngg^{-1}) at T0, T1, T2, T3, and T4 stages under N1, N2, and N3 rates, and yellow, green and blue colors represent different N rates, while the color shades reflected the level of the contents. Dash lines indicated the supposed pathway of auxin biosynthesis while the solid line showed the detected pathway of auxin biosynthesis. Lower case behind the numbers in the tables represented significant differences, $p \leq 0.05$.

and N3-T1. The contents of IAM involved in the IAM pathway first showed a decreasing trend in each FMd stage under each nitrogen application rate, and then went up. In terms of the IAM contents at each FM initiation stage, the value under N3 was significantly higher than N2 and N1, with increments of 105.5 and 265.9%. IAA was biosynthesized under the combined action of IAN and IAM, the contents of which were minimized at FMd under each nitrogen application rate. After FMd, the content of IAA then rose and reached the highest value at the T4 stage under three nitrogen application rates. Aimed at the FM initiation stage, IAA content at N3-T1 was 30.7% significantly lower than that at N1-T3, while showing no difference compared to N2-T2 (Figure 3).

Temporal changes of gene expressions involved in auxin biosynthesis, auxin signaling transduction, and FM differentiation

To further reveal the molecular mechanism of the regulation of advanced timing of the transition from vegetative to reproductive stage under a higher nitrogen application rate, the RNA-seq analysis was performed for samples with different FMd stages as mentioned previously. The analysis of the annotation, differentially expressed genes, and gene ontology was listed in Supplementary Tables S1–S3. However, the current study only focused on the auxin metabolism, therefore, the expression of the genes only correlated with auxin biosynthesis, signaling transduction, and FM differentiation was assayed, and all of them were listed in Supplementary Tables S4–S6.

In total, seven kinds of auxin biosynthesis genes were identified, which were *ASA1/ASA2*, *IGS*, *TSA1*, *TSB1*, *CYP79B*, *NIT1/NIT2*, and *AMI1*, and the general expression modes were shown in Figure 3. In the IAOx pathway, six copies of *CYP79B* gene were identified in this study, which is the key gene functioning on the conversion of Trp to IAOx. There was one copy (*BnaC03G0681300*) that showed downregulation mode throughout the FMd stages under all nitrogen application rates. However, the other five copies of *CYP79B* showed higher expression levels at the first three stages, with the expression level at N3 obviously higher than at N2 and N1, respectively, then declined rapidly. *NIT* was the second key gene in the IAOx pathway which functions on the conversion of IAN to IAA. There were five copies of *NIT* identified in the study. Among those genes, *BnaC02G0078200* showed a strong upregulation mode, and the expression level of which reached the highest

at each FM initiation stage under different N application rates. However, *BnaA02G0068000* showed a very weak upregulation mode and the remaining three copies showed downregulation mode under three nitrogen application rates. *AMI1* involves in the IAM pathway and functions in transferring IAM to IAA. There were two copies of *AMI1* found in the study. Both copies showed higher expression levels with upregulation mode at the first two stages (T0 and T1) and then also decreased to a relatively low level with downregulation mode (Figure 3). The expression of genes encoding key enzymes before the pathway of auxin biosynthesis including *ASA1/ASA2*, *IGS*, *TSA1,* and *TSB1* were also analyzed. Among those genes, *IGS* was not identified in this study. There were seven copies of *ASA1/ASA2* identified in this study. However, *BnaA03G0229800*, *BnaC03G0270600*, and *BnaC03G0270700* showed downregulation mode during the FMd stage under all nitrogen application rates. The remaining four copies showed strong expression under at T0 to T2 stages of FMd, while generally showing downregulation mode from the T3 stage. Furthermore, at the T0 stage, the highest expression level was found in N3 treatment. As for *TSA1*, there were four copies identified. Two of them (*BnaA09G0026900* and *BnaC09G0011700*) showed downregulation mode throughout the FMd stage under all nitrogen application rates. The other two copies, *BnaA09G0505300* and *BnaC08G0346300*, generally showed upregulation mode from T0 to T2 under three nitrogen application levels. Furthermore, both genes showed the highest expression levels at N3 in T0, T2, and T4 FMd stages (Figure 3).

There were totally five kinds of auxin signaling transduction genes identified, which were *TIR1*, *AUX/IAA*, *ARF*, *GH3,* and *SAUR*. *AUX/IAA* genes are a large family in plants and there were 47 genes found in the study. The genes can be divided into three types according to the expression levels. The first group was a strong expression with upregulation mode, which is composed of eight genes generally throughout the FMd stage under three nitrogen application rates. The second group was a weak expression with upregulation, which had 12 genes. And the third group was downregulation, which had 27 genes. Among the eight strongly expressed *AUX/IAA* genes, the expression levels at T1 and T3 were much stronger than in other stages. As for the nitrogen application rate, N2 had higher expression levels than N1 and N3 in general. A similar expression trend of the weak upregulation genes was observed. As for *TIR1*, there were seven genes found in the study. These genes had similar expression mode, which was the higher expression levels at T1 and T3 FMd stages. Furthermore, higher expression levels at N2 and N3 were found in all genes at the T1 FMd stage. However, at T2 and T3 FMd stages, the reverse trend, that is, higher expression levels of all genes at N1 was observed. At the T4 FMd stage, only *BnaA03G0267400* showed upregulation in all genes under three nitrogen application rates. As for *ARF*, there were two genes, *BnaA01G0267900* and *BnaC01G0328100*, that showed strong expression with upregulation, especially at T1 and T4 stages. Five *ARF* genes with downregulation mode were identified, however, *BnaA07G0137300*, *BnaA08G0250200*, and *BnaC05G0168800* had very strong downregulation mode in all FMd stages and nitrogen application rates. Generally, the N2 treatment had higher expression levels than in other nitrogen application rates. As for *GH3*, there were four genes identified in the study. Two genes, *BnaA03G0053200* and *BnaA09G0587900*, were in the downregulation mode nearly in all FMd stages and nitrogen application levels. One gene, *BnaC09G0017400*, had a strong expression with upregulation mode at most FMd stages. However, there was no obvious expression trend for responding to three nitrogen application rates at each differentiation stage. For *SAUR*, there were 16 genes identified in the study. Among them, seven genes showed upregulation mode at most FMd stages under three nitrogen application rates, and one gene, *BnaC01G0155600*, with relatively strong expression pattern. The gene had very strong expression levels at T2 and T3 stages under N3 treatment (Figure 4).

A total of 10 kinds of genes related to FM differentiation were identified, which were divided into two categories, which were signaling integration (*SOC1*, *FD*, *FT*, *FUL*, *TFL1*) and FM identity (*CAL*, *LMI1*, *LFY*, *AIL6*, *ANT*). As for *SOC1*, six copies were identified and two copies were in the downregulation mode throughout all FMd stages under three nitrogen application rates. The other four copies had an increasing trend of expression levels from T0 and higher expression levels at the T3 and T4 stages. As for *FD*, six copies were identified as well. Three copies were mainly in the downregulation mode during FMd stages and three nitrogen application rates. However, the other three copies were strongly expressed at the T3 stage with the highest expression amount at N2 treatment. One *FD* copy, *BnaA03G0552500*, showed a relatively higher expression amount with upregulation mode at T1, T2, and T4 in addition to T3. Furthermore, higher expression levels were found under N2 and N3 than that under N1 at T1 and T4. As for *FT*, there was only one copy detected in the study. The expression level was in the increasing trend during the FMd stage. At the T2 stage, the expression level of *FT* was in the upregulation mode at N1 while that was in the opposite mode at N2 and N3. However, the strongest expression of the gene was at N2 at T3. Although there were six copies identified for *FUL*, only one copy, *BnaA03G0404700*, showed upregulation mode. The highest expression level was found at the T3 stage and small differences in the expression level were found among three nitrogen application rates. For *TFL1*, there were four copies identified in the study, and two copies were in the downregulation mode throughout the FMd stage and all nitrogen application rates. Two copies, *BnaA03G0012400* and *BnaC03G0016500*, showed very strong expression at the T3 stage. Furthermore, *BnaA03G0012400* had the highest expression amount at T1 to T3 at N3. In the FM identity gene systems, there were six and four copies identified in the study for *CAL* and *LFY*, respectively.

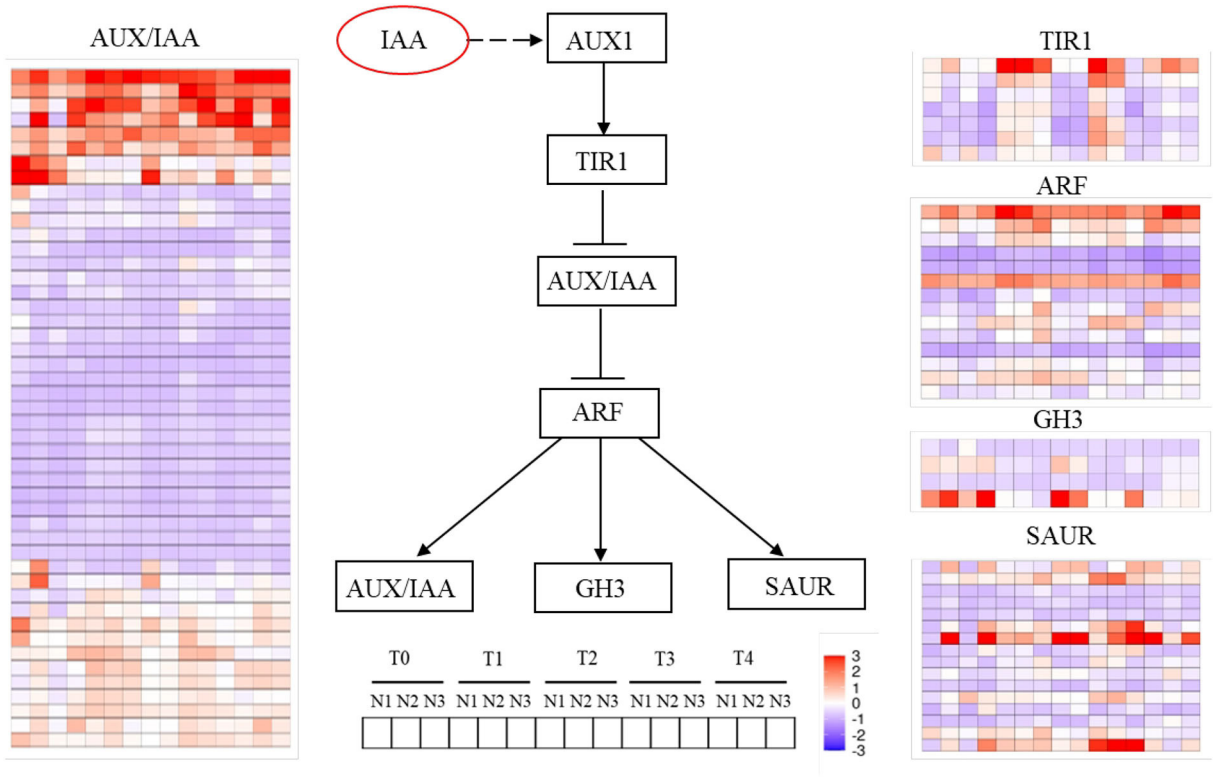


FIGURE 4
Changes of genes involved in auxin signaling transduction at T0, T1, T2, T3, and T4 sampling stages under low N (N1), normal N (N2), and high N (N3) rates. *ARF, Auxin response factor; AUX1, Auxin influx carrier 1; AUX/IAA, Auxin responsive protein; TIR1, Transporter inhibitor response 1*. Color bars represented the trends of expression levels after centralization and standardization by the R programming language and red color represents higher expression, while purple represents lower expression relatively.

However, only one copy of both genes, *BnaC02G033300 (CAL)* and *BnaC03G0561300 (LFY)*, respectively, showed strong expression with up-regulation mode at all FMd stages and nitrogen application rates. For *BnaC02G0233300* (*CAL*), the expression level at T3 was the highest followed by T1. However, the expression levels at N1 and N2 were much higher than that of N3 at the T3 stage, while the reverse trend was found in T1. For *LFY*, the expression level was increased from T0 to T3. The highest expression level of *LFY* was at T3 and N2 treatment. There were six copies identified for *LMI1*, *AIL6*, and *ANT*, and all of them showed three copies mainly having upregulation mode during FMd stages and three nitrogen application rates. For *BnaA10G0288900* (*LMI1*), the highest expression level was found at T2 with N2 treatment which was in accordance with the timing of FMd at this stage. As for *BnaC09G0552500* (*AIL6*), there were no relatively small differences in expression levels among FMd stages and nitrogen application rates. As for *BnaA01G0012600*, *BnaA03G0565600*, and *BnaC07G0540500* (*ANT*), all of them showed the highest expression levels at T1 and N3, T2 and N1, and T3 and N2 treatments, respectively (Figure 5).

Discussion

Accelerating the initiation of FM differentiation by a higher nitrogen application rate is an effective way for yield increasing

FM differentiation is the beginning of the transition from vegetative to reproductive growth of flowering plants. The initiation time and differentiation ability of FM determine the number of flower buds formed, which is of great significance to the establishment of rapeseed yield (Wrucke et al., 2019). The initiation of FM differentiation occurs at the SAM, and the earlier FM initiation time will enhance the efficiency of rapeseed utilizing the late autumn temperature and light resources, which can promote reproductive growth a lot (Fiebelkorn and Rahman, 2016).

In this research, we found that the initiation time of FM was largely regulated by different based-N-fertilizer levels. The FM initiation time was at December 7th under low N level while with increase of N fertilization, the FMd could be promoted

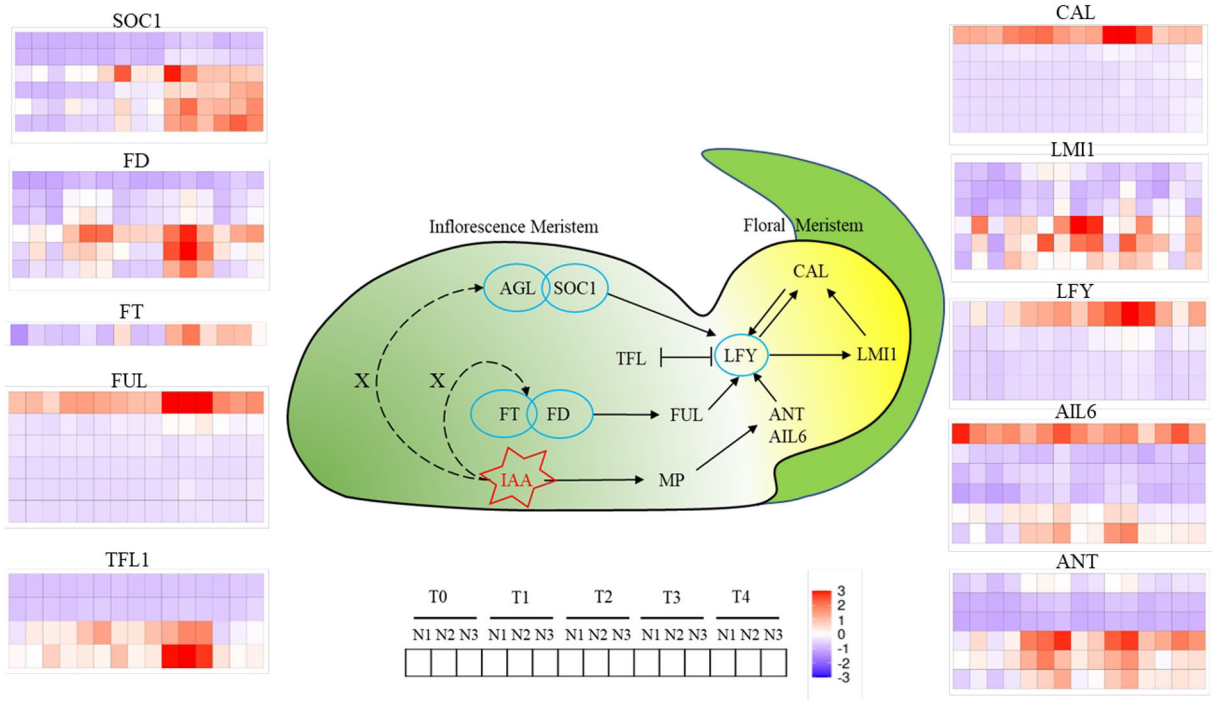


FIGURE 5
Changes of genes involved in floral meristems identity genes at T0, T1, T2, T3, and T4 stages under low N (N1), formal N (N2), and high N (N3) rates. Color bars represented the trends of expression levels after centralization and standardization by R programming language and red color represents higher expression, while purple represents lower expression relatively.

by 3 to 7 days. The advanced dates can strive temperature and light resources of late-autum a lot for rapeseed growth and hence the increase of siliques and yield up to 22.6 and 50.5%, respectively (Figure 1). To determine if the increase in yield was related to the initiation and development of FM, a scanning electron microscope was utilized for the observation and identification of FM morphology. The results showed that the appearance of floral meristems, sepal primordium, gynoecium, and androecium primordium occurred earlier in higher N rates than in lower. What's more, the emergence of FMs enlarging to ball-shaped structures and the separation of FMs from IM also went earlier (Figure 2), which accelerated the establishment of reproductive organs, hence, more temperature and lights were utilized.

Trp-IAOx-IAN-IAA biosynthesis pathway might be the major auxin biosynthetic pathway in rapeseed SAM, and the decrease of endogenous auxin contents may be the essential condition for FM initiation

FM differentiation, as the transition from SAM to floral meristem, is induced by a series of environmental signal pathways, including photoperiod, low temperature, and carbohydrates, and is accompanied by a large number of physiological, biochemical, and gene expression changes. Studies have shown that plants have already started the regulation and transformation of the physiological and molecular mechanisms before the morphological change of SAM (Liu et al., 2009). Therefore, illuminating the changes in physiology, biochemistry, and gene expression in different processes before and after FM differentiation is of great significance for explaining the mechanism of FM differentiation, shortening the growth period, and improving the yield of rapeseed (Yan et al., 2021).

The homeostasis of endogenous hormones, as the basis of plant vegetative growth and reproductive growth, plays an important role in the flowering process of plants and is also one of the important conditions for flowering induction (Daviere and Achard, 2016; Castorina and Consonni, 2020). Auxin as a major phytohormone controls numerous aspects of rapeseed development and coordinates the responses to the environment, and so does the initiation of FM and organs development (Brumos et al., 2018). However, the role of how auxin determines the development and transition of IM to FM remains less mechanistically. According to previous studies in *Arabidopsis*, auxin was accumulated at the position of FM and gradually decreased with the extension of distance, and it is the gradient that contributes to the development and

differentiation of floral organs (Benková et al., 2003). However, the changes in auxin contents in SAM and the gene expression in its biosynthetic pathway in different periods are not clear. In this study, we identified the contents of auxins at five stages under three different N levels of SAM through UPLC-MS/MS, and four auxins were detected, which involved two auxin biosynthetic ways, which were IAOx way and IAM way, respectively. Previous research showed that the indole pyruvic acid (IPA) way is the main auxin biosynthetic way, while IAOx and IAM contribute less (Mashiguchi et al., 2011; Won et al., 2011). However, we detected a large amount of auxin biosynthetic precursors in IAOx and IAM ways, which were IAN and IAM, and the contents of IAN were 1,740 times of IAM in terms of the maximum for each other, while no auxins involved in IPA way were detected, which indicated that Trp-IAOx-IAN-IAA might be the major auxin biosynthetic pathway in rapeseed shoot apical meristems. What's more, we found that both of the contents of Trp and IAA showed a trend of increase in the first and then a sharp decrease in the FM initiation stage followed by a gradual rise (Figure 3). Through the combination of the images of SEM, we presumed that IM needs relatively low auxin contents as the signal to initiate the differentiation of FM, then accumulated more auxin contents for the following up development of FMs, and so did the sepal primordium. At the same time, we found that the expression of genes such as *TSA1*, *TSB1*, *CYP79B*, *NIT,* and *AMI1* was higher expressed in 1 to 2 stages earlier than each FM initiation stage, while the maximum value came earlier under high N level than under low N level, which indicated that high N rate accelerates the activity of auxin biosynthesis genes, while this acceleration occurred 1 to 2 stages before FM initiation.

Auxin signaling transduction and FM identity genes were largely regulated under different nitrogen levels

Interactions of endogenous growth and developmental signals with gene expressions constitute a complex pathway for plant reproductive growth induction. We identified that most of the auxin signaling transduction genes were upregulated in T1 to T3 stages, relative to one stage later than auxin biosynthesis, while the higher expression of FM differentiation-related genes came at the last two stages.

Among them, *AGL24*, *SOC1*, *FT*, *FD,* and *FUL* functioned as integrating environmental and other signals in SAM and transferring the signal to *LFY* for *FM* identify and specification (Liu et al., 2009). SOC1 is highly expressed in IMs and induces the expression of LFY by binding to the promoter region of *LFY* through the AGL24-SOC1 protein complex (Li et al., 2008). FT-FD complex also promotes the activity of SOC1 or FUL in IMs, hence the expression of IM identity genes such as LFY and CAL were enhanced (Li et al., 2008; Liu et al., 2008). According to our research, the expression of *SOC1* and *FD* began to rise at the T1 stage under the N3 level, while the significant upregulation occurred 1 to 2 stages later than N3 under N1 and N2 levels (Figure 5). Terminal flower 1 (*TFL1*) as a repressor of flowering, is higher expressed in SAM, while less expressed in FM, which functions by maintaining the disordered growth of IM and inhibiting the expression of *LFY* in IM (Schiessl et al., 2017; Zhang et al., 2018). In this study, we can find that, under normal or high N rate, the expression of *TFL1* was repressed until the T4 stage, while the upregulation occurred two stages earlier under a low N rate compared with N3, which indicated that the inhibition of *TFL1* on FM development was weakened under high N rate. *LFY*, as a central gene of FM, is reported to regulate the formation of FM through auxin signaling transduction. *ANT* and *AIL6* as the upstream genes of *LFY* functioned parallel and redundantly to regulate the expression of *LFY* and the response to auxins (Yamaguchi et al., 2016). According to our results, *ANT* and *AIL6* were upregulated at the T1 stage, 1 to 2 stages earlier than LFY, which was consistent with the previous research (Yamaguchi et al., 2016). At the same time, *CAL*, another distinctive FM-identified gene, was regulated by *LFY* directly or through *LMI1*, and the expression was relatively higher and earlier under a high N rate.

In brief, FM-identified genes were higher and earlier expressed under high N rate compared with normal or low N rate, while the activity of FM repressor was limited, which co-contribute to the advanced development of FM under high N rate.

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://www.ncbi.nlm.nih.gov/ - PRJNA846178.

Author contributions

SH designed the experiment and revised the manuscript. PH investigated (all) and wrote the manuscript. BL analyzed the RNA-seq data. HH analyzed the auxin content measurement. BX and LH collected the agronomic data and microscopy imaging. All authors approved the final version of the article.

Funding

This work was supported by the National Key Research and Development Project (2018YFD1000900), the Zhejiang

Key Laboratory of Digital Dry Land Crops (2022E10012), and the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02064).

Acknowledgments

Great appreciation are given to the editor and reviewers' critical comments on the improvement of the 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.2022.927662/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The growth condition of mean temperature and rainfall during the field trials.

SUPPLEMENTARY FIGURE S2
Correlation analysis between ratio of floral meristems differentiation, siliques per plant, and seed yield per acre. The values above the double-sided arrows represent the correlation coefficients on 30th Nov, 04th Dec, and 07th Dec, respectively.

SUPPLEMENTARY TABLE S1
Total differentiation expressed genes and GO, KEGG analyzed genes in the study with different nitrogen application rates and floral meristem differentiation timings.

SUPPLEMENTARY TABLE S2
Total upregulated and downregulated genes in the study with different nitrogen application rates and floral meristem differentiation timings.

SUPPLEMENTARY TABLE S3
Annotation of total differentiation expressed genes.

SUPPLEMENTARY TABLE S4
Genes were identified in correlation to auxin biosynthesis.

SUPPLEMENTARY TABLE S5
Genes were identified in correlation to auxin signaling transduction.

SUPPLEMENTARY TABLE S6
Genes were identified in correlation to floral meristem differentiation.

References

Abley, K., Sauret-Gueto, S., Maree, A. F. M., and Coen, E. (2016). Formation of polarity convergences underlying shoot outgrowths. *Elife* 5, e18165. doi: 10.7554/eLife.18165.053

Balkunde, R., Kitagawa, M., Xu, X. M., Wang, J., and Jackson, D. (2017). SHOOT MERISTEMLESS trafficking controls axillary meristem formation, meristem size and organ boundaries in *Arabidopsis*. *Plant J.* 90, 435–446. doi: 10.1111/tpj.13504

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertov,á, D., Jürgens, G., et al. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* 115, 591–602. doi: 10.1016/S0092-8674(03)00924-3

Bilsborough, G. D., Runions, A., Barkoulas, M., Jenkins, H. W., Hasson, A., Galinha, C., et al. (2011). Model for the regulation of *Arabidopsis thaliana* leaf margin development. *Proc. Natl. Acad. Sci. USA* 108, 3424–3429. doi: 10.1073/pnas.1015162108

Brumos, J., Robles, L. M., Yun, J., Yu, T., Jackson, S., Alonso, J. M., et al. (2018). Local auxin biosynthesis is a key regulator of plant development. *Dev. Cell* 47, 306–318. doi: 10.1016/j.devcel.2018.09.022

Carles, C. C., and Fletcher, J. C. (2003). Shoot apical meristem maintenance: the art of a dynamic balance. *Trends Plant Sci.* 8, 394–401. doi: 10.1016/S1360-1385(03)00164-X

Castorina, G., and Consonni, G. (2020). The role of brassinosteroids in controlling plant height in *Poaceae*: A genetic perspective. *Int. J. Mol. Sci.* 21, 16. doi: 10.3390/ijms21041191

Daviere, J. M., and Achard, P. (2016). A pivotal role of DELLAs in regulating multiple hormone signals. *Mol. Plant* 9, 10–20. doi: 10.1016/j.molp.2015.09.011

Fiebelkorn, D., and Rahman, M. (2016). Development of a protocol for frost-tolerance evaluation in rapeseed/canola (*Brassica napus* L.). *Crop J.* 4, 147–152. doi: 10.1016/j.cj.2015.11.004

Fu, W. Q., Chen, D. Z., Pan, Q., Li, F. F., Zhao, Z. G., Ge, X. H., et al. (2018). Production of red-flowered oilseed rape via the ectopic expression of *Orychophragmus violaceus OvPAP2*. *Plant Biotechnol. J.* 16, 367–380. doi: 10.1111/pbi.12777

Fu, Y. F., Zhang, Z. W., Yang, X. Y., Wang, C. Q., Lan, T., Tang, X. Y., et al. (2020). Nitrate reductase is a key enzyme responsible for nitrogen-regulated auxin accumulation in *Arabidopsis* roots. *Biochem. Bioph. Res. Comm.* 532, 633–639. doi: 10.1016/j.bbrc.2020.08.057

Gu, P. Y., Luo, F. F., Tao, E. Q., Li, Y., Wang, D. J., Wu, X., et al. (2022). Higher nitrogen content and auxin export from rice tiller enhance low-ammonium-dependent tiller outgrowth. *J. Plant Physiol.* 268, 153562. doi: 10.1016/j.jplph.2021.153562

Hamada, S., Onouchi, H., Tanaka, H., Kudo, M., Liu, Y., Shibata, D., et al. (2000). Mutations in the *WUSCHEL* gene of *Arabidopsis thaliana* result in the development of shoots without juvenile leaves. *Plant J.* 24, 91–101. doi: 10.1046/j.1365-313x.2000.00858.x

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., et al. (2008). A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Dev. Cell* 15, 110–120. doi: 10.1016/j.devcel.2008.05.002

Li, Y., Zhou, C., Yan, X., Zhang, J., and Xu, J. (2016). Simultaneous analysis of ten phytohormones in *Sargassum horneri* by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry. *J. Sep. Sci.* 39, 1804–1813. doi: 10.1002/jssc.201501239

Liu, C., Chen, H., Er, H. L., Soo, H. M., Kumar, P. P., Han, J. H., et al. (2008). Direct interaction of *AGL24* and *SOC1* integrates flowering signals in *Arabidopsis*. *Development* 135, 1481–1491. doi: 10.1242/dev.020255

Liu, C., Thong, Z., and Yu, H. (2009). Coming into bloom: the specification of floral meristems. *Development* 136, 3379–3391. doi: 10.1242/dev.033076

Liu, Y. J., Ye, S. H., Yuan, G. G., Ma, X. W., Heng, S. P., Yi, B., et al. (2020). Gene silencing of *BnaA09.ZEP* and *BnaC09.ZEP* confers orange color in *Brassica napus* flowers. *Plant J.* 104, 932–949. doi: 10.1111/tpj.14970

Luo, L. J., Zeng, J., Wu, H. J., Tian, Z. X., and Zhao, Z. (2018). A molecular framework for auxin-controlled homeostasis of shoot stem cells in *Arabidopsis*. *Mol. Plant* 11, 899–913. doi: 10.1016/j.molp.2018.04.006

Mashiguchi, K., Tanaka, K., Sakai, T., Sugawara, S., Kawaide, H., Natsume, M., et al. (2011). The main auxin biosynthesis pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 108, 18512–18517. doi: 10.1073/pnas.1108434108

Narnoliya, L., Basu, U., Bajaj, D., Malik, N., Thakro, V., Daware, A., et al. (2019). Transcriptional signatures modulating shoot apical meristem morphometric and plant architectural traits enhance yield and productivity in chickpea. *Plant J.* 98, 864–883. doi: 10.1111/tpj.14284

Nidhi, S., Preciado, J., and Tie, L. (2021). Knox homologs shoot meristemless (STM) and KNAT6 are epistatic to CLAVATA3 (CLV3) during shoot meristem development in *Arabidopsis thaliana*. *Mol. Biol. Rep.* 48, 6291–6302. doi: 10.1007/s11033-021-06622-4

Olas, J. J., Van Dingenen, J., Abel, C., Działo, M. A., Feil, R., Krapp, A., et al. (2019). Nitrate acts as the *Arabidopsis thaliana* shoot apical meristem to regulate flowering time. *New Phytol.* 223, 814–827. doi: 10.1111/nph.15812

Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. (2017). Flowering time gene variation in Brassica species shows evolutionary principles. *Front. Plant Sci.* 8, 1742. doi: 10.3389/fpls.2017.01742

Shi, B. H., and Vernoux, T. (2022). Hormonal control of cell identity and growth in the shoot apical meristem. *Curr. Opin. Plant Biol.* 65, 102111. doi: 10.1016/j.pbi.2021.102111

Šimura, J., Antoniadi, I., Široká, J., Tarkowská, D., Strnad, M., Ljung, K., et al. (2018). Plant Hormonomics: Multiple Phytohormone Profiling by Targeted Metabolomics. *Plant Physiol.* 177, 476–489. doi: 10.1104/pp.18.00293

Tooke, F., and Battey, N. (2003). Models of shoot apical meristem function. *New Phytol.* 159, 37–52. doi: 10.1046/j.1469-8137.2003.00803.x

Traas, J. (2018). Molecular networks regulating meristem homeostasis. *Mol. Plant* 11, 883–885. doi: 10.1016/j.molp.2018.06.004

Vernoux, T., Besnard, F., and Traas, J. (2010). Auxin at the shoot apical meristem. *CSH. Perspect. Biol.* 2, a001487. doi: 10.1101/cshperspect.a001487

Vernoux, T., Kronenberger, J., Grandjean, O., Laufs, P., and Traas, J. (2000). *PIN-FORMED 1* regulates cell fate at the periphery of the shoot apical meristem. *Development* 127, 5157–5165. doi: 10.1242/dev.127.23.5157

Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., et al. (2011). Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF *ARABIDOPSIS* and YUCCAs in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 108, 18518–18523. doi: 10.1073/pnas.1108436108

Wrucke, D. F., Talukder, Z. I., Rahman, M., and Snowdon, R. (2019). Genome-wide association study for frost tolerance in rapeseed/canola (*Brassica napus* L.) under simulating freezing conditions. *Plant Breed.* 139, 356–367. doi: 10.1111/pbr.12771

Xin, W., Wang, Z. C., Liang, Y., Wang, Y. H., and Hu, Y. X. (2017). Dynamic expression reveals a two-step patterning of *WUS* and *CLV3* during axillary shoot meristem formation in *Arabidopsis*. *J. Plant Physiol.* 214, 1–6. doi: 10.1016/j.jplph.2017.03.017

Yamaguchi, N., Jeong, C. W., Nole-Wilson, S., Krizek, B., and Wager, D. (2016). AINTEGUMENTA and AINTEGUMENTA-LIKE6/PLETHORA3 induce *LEAFY* expression in response to auxin to promote the onset of flower formation in *Arabidopsis*. *Plant Physiol.* 170, 283–293. doi: 10.1104/pp.15.00969

Yan, G., Yu, P., Tian, X., Guo, L., and Dai, C. (2021). DELLA Proteins BnaA6.RGA and BnaC7.RGA negatively regulate fatty acid biosynthesis by interacting with BnaLEC1s in *Brassica napus*. *Plant Biotechnol. J.* 19, 2011–2026. doi: 10.1111/pbi.13628

Zhang, Y., Zhang, D., Yu, H., Lin, B., Fu, Y., and Hua, S. (2016). Floral initiation in response to planting date reveals the key role of floral meristem differentiation prior to budding in canola (*Brassica napus* L.). *Front. Plant Sci.* 7, 1369. doi: 10.3389/fpls.2016.01369

Zhang, Y., Zhang, D., Yu, H., Lin, B., Hua, S., Ding, H., et al. (2018). Location and mapping of the determinate growth habit of *Brassica napus* by bulked segregant analysis (BSA) using whole genome re-sequencing. *Sci. Agric. Sinica* 51, 3029–3039. doi: 10.3864/j.issn.0578-1752.2018.16.001

Zhao, Y. D. (2011). 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

Zhou, Y. Y., Honda, M., Zhu, H. L., Zhang, Z. H., Guo, X. W., Li, T. H., et al. (2015). Spatiotemporal sequestration of miR165/166 by *Arabidopsis* Argonaute10 promotes shoot apical meristem maintenance. *Cell Rep.* 10, 1819–1827. doi: 10.1016/j.celrep.2015.02.047

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fpls.2022.962592

OPEN ACCESS

EDITED BY
Sarvajeet Singh Gill,
Maharshi Dayanand University, India

REVIEWED BY
Mingli Yan,
Hunan Academy of Agricultural Sciences, China
Wei Qian,
Southwest University, China

*CORRESPONDENCE
Xiaohui Cheng
chengxiaohui@caas.cn
Junyan Huang
huangjy@oilcrops.cn

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

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

RECEIVED 06 June 2022
ACCEPTED 23 August 2022
PUBLISHED 14 September 2022

CITATION
Liu J, Wu Y, Cui X, Zhang X, Xie M, Liu L, Liu Y, Huang J, Cheng X and Liu S (2022) Genome-wide characterization of ovate family protein gene family associated with number of seeds per silique in *Brassica napus*. *Front. Plant Sci.* 13:962592. doi: 10.3389/fpls.2022.962592

Genome-wide characterization of ovate family protein gene family associated with number of seeds per silique in *Brassica napus*

Jie Liu†, Yupo Wu†, Xiaobo Cui, Xiong Zhang, Meili Xie, Lijiang Liu, Yueying Liu, Junyan Huang*, Xiaohui Cheng* and Shengyi Liu

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

Ovate family proteins (OFPs) were firstly identified in tomato as proteins controlling the pear shape of the fruit. Subsequent studies have successively proved that OFPs are a class of negative regulators of plant development, and are involved in the regulation of complex traits in different plants. However, there has been no report about the functions of OFPs in rapeseed growth to date. Here, we identified the OFPs in rapeseed at the genomic level. As a result, a total of 67 members were obtained. We then analyzed the evolution from *Arabidopsis thaliana* to *Brassica napus*, illustrated their phylogenetic and syntenic relationships, and compared the gene structure and conserved domains between different copies. We also analyzed their expression patterns in rapeseed, and found significant differences in the expression of different members and in different tissues. Additionally, we performed a GWAS for the number of seeds per silique (NSPS) in a rapeseed population consisting of 204 natural accessions, and identified a new gene *BnOFP13_2* significantly associated with NSPS, which was identified as a novel function of OFPs. Haplotype analysis revealed that the accessions with haplotype 3 had a higher NSPS than other accessions, suggesting that *BnOFP13_2* is associated with NSPS. Transcript profiling during the five stages of silique development demonstrated that *BnOFP13_2* negatively regulates NSPS. These findings provide evidence for functional diversity of OFP gene family and important implications for oilseed rape breeding.

KEYWORDS

ovate family protein, *Brassica napus*, GWAS, yield traits, negative regulator, silique development

Introduction

Oilseed crops are both an important source of edible vegetable oil and a valuable material of animal feed (Felten et al., 2013; D'Avino et al., 2015). As the second largest source of edible vegetable oil, *Brassica napus* (*B. napus*) provides about 13–16% of the total edible vegetable oil for the world (Wang et al., 2018), and the market demand for it is still increasing in recents. Therefore, great efforts have been made to improve its yield. Plant architecture, including main inflorescence length, plant height and branch number, indirectly affects the oilseed rape yield (Reinhardt and Kuhlemeier, 2002; Wang and Li, 2008; Li et al., 2019a), while siliques per plant (SPP), number of seeds per silique (NSPS), and total seed weight (TSW) directly determine the plant yield (Clarke and Simpson, 1978). Similarly, in *Arabidopsis*, the silique is closely associated with the final yield, and silique-related traits such as silique length (SL) and silique volume (SV) affect the morphology and photosynthetic substances, thus influencing the yield (Ferrándiz et al., 1999). In siliques, NSPS and SL are more important traits. In recent years, many QTLs controlling SL and NSPS have been identified at almost all chromosomes through QTL mapping and genome wide association study (GWAS) (Fu et al., 2015; Wang et al., 2016; Yang et al., 2016, 2017; Luo et al., 2017; Zhu et al., 2020). To date, two genes controlling the SL, *BnaA9.ARF18* and *BnaA9.CYP78A9*, have been cloned in *B. napus*, which affect SL by regulating cell elongation of the silique wall (Liu et al., 2015a; Shi et al., 2019). Additionally, *BnaC9.SMG7B* has been cloned as a positive regulator of NSPS, which regulates the formation of normal female gametophyte and finally determines the formation of mature ovules (Li et al., 2015). In recent studies, *BnaA08g07940D* and *BnaA08g07950D* were identified as putative candidate genes of a major QTL controlling NSPS by fine mapping (Jiao et al., 2021). Besides, in *Brassica juncea*, *BjCLV1* was found to affect NSPS through the formation of trilocular siliques (Wang et al., 2021). In *Arabidopsis*, cytokinin and brassinosteroid were found work coordinately to promote ovule initiation and then increase NSPS (Zu et al., 2022). Certainly, the yield is also affected by diseases such as Sclerotinia stem rot, clubroot, blackleg disease and stem canker. Fortunately, many resistance QTLs have been identified. Therefore, integration of these elite QTL alleles controlling different traits into elite cultivars with better plant architecture may be a promising strategy to improve the yield of oilseed rape.

Domestication of fruit-bearing crops involves long-term artificial selection from various wild plant species, and a significant hallmark in this process is the explosive increase in fruit shape variations (Williams, 1965; Hopping et al., 1986). The recessive locus ovate controlling the pear shape of fruit and elongated fruit shape in tomato was identified about one hundred years ago, but the gene was cloned until 20 years ago (Price and Drinkard, 1908; Liu et al., 2002). The ovate proteins were identified as a class of negative regulators in plant development, which contain a C-terminal conserved domain and Von Willebrand factor type C domain, which are conserved in tomato, rice and *Arabidopsis* (Liu et al., 2002; Wang et al., 2011). Subsequently, studies of plant ovate family proteins (OFPs) have been successively reported. Researchers are increasingly aware of their functions though the exact mechanism remains poorly understood. OFPs play important roles in plant growth and development, and their functions have been studied in both model plants and crops. In *Arabidopsis*, *AtOFP5* affects the cell-fate switch of synergid to egg cell in mature embryo sacs by suppressing the ectopic activity of BELL-KNOX TALE complex (Pagnussat et al., 2007); *AtOFP4* and *AtOFP1* are involved in secondary cell wall formation; and the *Atofp4* mutant exhibited thicker interfascicular fiber cell wall and thinner xylary fiber cell wall (Li et al., 2011; Wang et al., 2011). AtOFP1 interacts with ATH1 to regulate flowering time and stem growth in *Arabidopsis* (Zhang et al., 2018). Besides, *AtOFP1* is involved in the development of male gamete and pollen activity as well as DNA repair (Hackbusch et al., 2005; Wang et al., 2010). In rice (*Oryza sativa*), overexpression of *OsOFP2* led to a series of variations in plant height, leaf morphology, seed shape and abnormity of vascular bundles in stems; *OsOFP2* suppresses the expression of *GA20ox* by modulating the function of KNOX-BELL and inhibits lignin biosynthesis, thereby affecting vasculature development (Schmitz et al., 2015). In radish (*Raphanus sativus*), *RsOFP2.3* is negatively associated with tuberous root elongation and the tuberous root shape (Wang et al., 2020). In peach (*Prunus persica*), *PpOFP1* regulates fruit shape (Cirilli and Rossini, 2021), and another similar study showed that PpOFP1 physically interacts with a ZF-HD_dimer domain protein PpZFHD1 and regulates the salt tolerance of tomato (Tan et al., 2021). In cotton (*Gossypium hirsutum*), *GhOFP4* was found to regulate fiber development (Gong et al., 2014). In *Capsicum annuum*, *CaOFP1* is involved in fruit shaping, and its different expression profiles would result in different shapes *via* negatively affecting the expression of *CaGA20ox1* (Tsaballa et al., 2011), and gene silencing of *CaOFP20* increased the fruit length (Borovsky et al., 2022). Ectopic expression of *CsOFP12-16c* from cucumber (*Cucumis sativus*) in *Arabidopsis* affects the silique development and causes blunt and shorter siliques (Han et al., 2022). Previous studies have also suggested that *CmOFP13* may control the fruit shape in melon (*Cucumis melo*) (Monforte et al., 2014; Ma et al., 2021). Also, the expression of *MaOFP1* was reported to be negatively associated with fruit ripening in banana (*Musa paradisiaca*) (Liu et al., 2015b). Although more and more functions of OFPs have been reported in many crops, there are still numerous unknown features remaining to be discovered.

B. napus (2n = 38, AACC) is a polyphyletic polyploidy formed by *B. oleracea* and *B. rapa* (Nagaharu, 1935; Allender and King, 2010), and has experienced whole genome duplication (WGD) (Jiao et al., 2011; Chalhoub et al., 2014). *Brassica napus napus* still retains two sets of chromosomes corresponding

to *B. oleracea* and *B. rapa* (Chalhoub et al., 2014). Hence, there are many duplicate genes from two sub-genomes or even the earlier progenitor *Arabidopsis*. Generally, *Arabidopsis* can serve as an efficient model plant in functional gene research. *AtOFPs* have been found to have different functions. However, multiple copies of *OFP* genes have been rarely reported in its closely related species *B. napus*. In this study, we identified the *BnOFP* gene family at the genomic level, analyzed its evolution from *Arabidopsis* to *B. napus* and compared the difference between copies. A new locus *BnOFP13_2* significantly correlated with NSPS was identified, which can be considered as a novel function of *OFPs*. RNA profiling during silique development suggested that *BnOFP13_2* negatively regulates NSPS. Our findings provide evidence for the functional diversity of OFP gene family and important implications for oilseed rape breeding.

Results

Identification and chromosomal distribution of ovate family proteins gene family in *Brassica napus*

By using the reported 20 AtOFP or AtOFP-like protein sequences as queries, a total of 67 *OFP* genes were identified through BLAST in *B. napus* and *Arabidopsis* databases, which were renamed according to their orthologous genes in *Arabidopsis*, and their physical and chemical properties were analyzed (**Supplementary Table 1**). Among the 20 *AtOFP* genes, *AtOFP6* and *AtOFP9* had no orthologous gene in *B. napus*; *AtOFP1* had only one orthologous gene *BnOFP1_1*; while *AtOFP2*, *AtOFP3* and *AtOFP5* all had six orthologous genes. The other *13 AtOFP* genes, respectively, had two to five orthologous genes in *B. napus* (**Supplementary Table 1**). The 67 *BnOFP* genes were unevenly distributed on the 20 chromosomes of A sub-genome (34 *BnOFP* genes) and C sub-genome (33 *BnOFP* genes) (**Figure 1**). In the A sub-genome, A08 chromosome had no *BnOFP* gene; chromosome A06 contained one *BnOFP* gene; and A02, A09, and A10 chromosomes, respectively, harbored five *BnOFP* genes. On the same chromosome, some *BnOFP* genes were closely located, while some other genes were far away from each other. For instance, *BnOFP17_3* and *BnOFP2_5* were closely located on chromosome A04, while *BnOFP7_4* and *BnOFP14_4* were located on both ends of chromosome A07. In the C sub-genome, chromosome C06 contained no *BnOFP* member, and eight *BnOFP* members were located on unknown chromosomes, while other chromosomes, respectively, harbored two to five members (**Figure 1**). These results suggested the occurrence of genome rearrangement and gene loss during polyploidization.

Phylogenetic and syntenic relationship of ovate family proteins gene family in *Brassica napus*

Phylogenetic relationship is an important indicator for gene functional research. The 87 OFP proteins from *B. napus* and *Arabidopsis* were used to construct a phylogenetic tree. As a result, five groups were clustered based on the sequence alignment (**Figure 2** and **Supplementary Table 2**). Group 2 was the largest clade (24 proteins), followed by group 5 (19 proteins), group 4 (17 proteins), group 3 (16 proteins), and group 1 (11 proteins). Among the five groups, group 2 and group 3 contained some *AtOFPs* and all their corresponding *B. napus* orthologous genes. For example, group 2 included *AtOFP5*, *AtOFP11*, *AtOFP12*, *AtOFP15*, *AtOFP16*, *AtOFP18* and *BnOFP5*, *BnOFP11*, *BnOFP12*, *BnOFP15*, *BnOFP16*, and *BnOFP18*. On the contrary, some *AtOFPs* and their orthologous *BnOFPs* were not clustered together. For example, *AtOFP10* was clustered in group 1, but the *BnOFP10* orthologs were clustered in group 4 or group 5. Genes with similar functions tended to be clustered in the same group. For instance, *AtOFP1* and *AtOFP4* were both involved in secondary cell wall formation and thus clustered in group 3 (Li et al., 2011; Wang et al., 2011); *AtOFP17* and *AtOFP20* were both paralogous genes from segmental duplication blocks, and were also clustered in the same group (Liu et al., 2014; **Figure 2**). During the polyploidization from *Arabidopsis* to *B. napus*, members in different groups had undergone loss-of-function and function divergence. Therefore, the *BnOFP* gene family members may participate in different biological processes in plant development.

Since *B. napus* is of polyphyletic polyploidy with syntenic relationship between the sub-genomes, we analyzed the syntenic relationship between the *BnOFPs* in A and C sub-genomes (**Figure 3**). We analyzed the *53 BnOFPs* with specific chromosomal locations, and finally identified 19 pairs of syntenic genes. No syntenic gene was detected on chromosome A07 and A08 as well as on C06 and C07, while other chromosomes contained one to five syntenic genes. Chromosome A10 included the most syntenic genes (five genes), with two on chromosome C05 and three on C09. Pairs of syntenic genes were in the same subfamily, such as *BnOFP7_3* and *BnOFP7_5*, and *BnOFP2_2* and *BnOFP2_5* (**Figure 3**). Gene phylogenetic and syntenic relationship can be used to explore the functions of unknown genes. Hence, our results may be of great significance for gene functional research in *B. napus*.

Gene structure and conserved domains of *BnOFP* genes

Gene structure was analyzed to investigate whether there are differences among the 67 *BnOFP* genes. As expected, large differences were found in gene length and exon number among

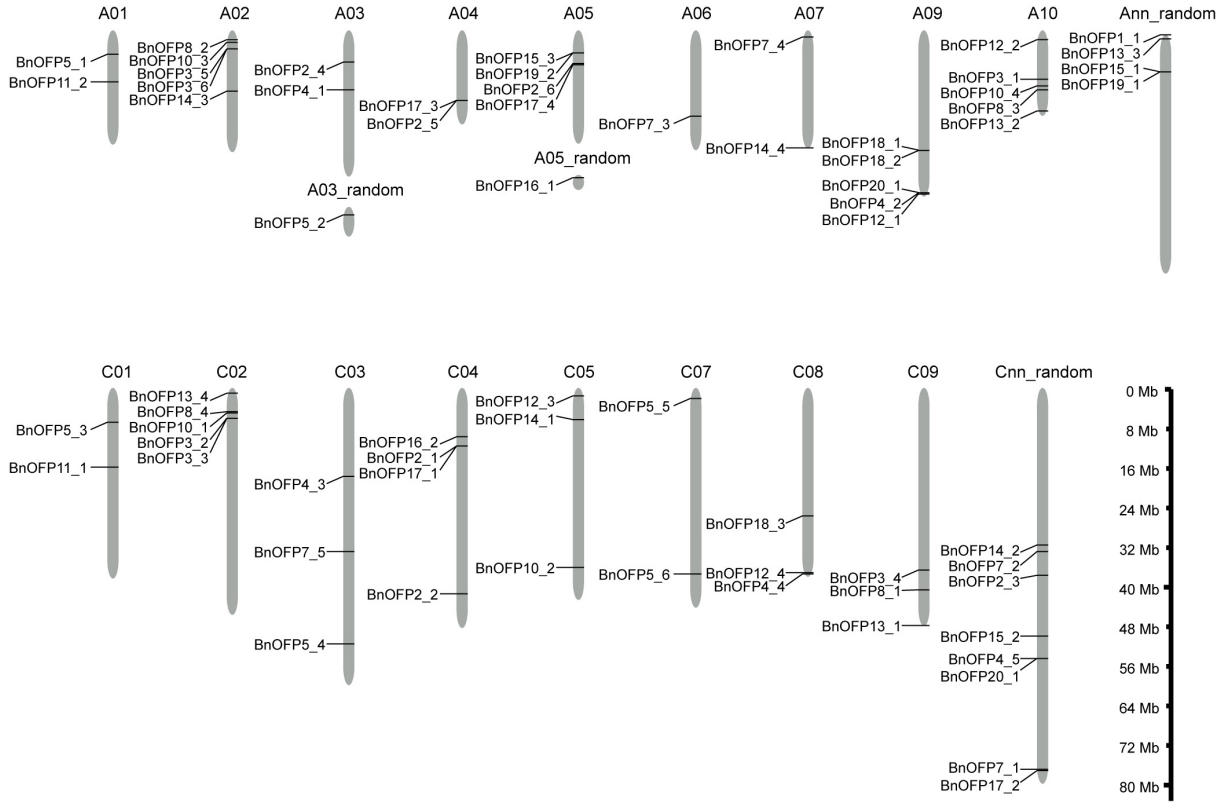


FIGURE 1
Chromosomal locations of *BnOFPs*. The 67 *BnOFP* genes were located on different chromosomes of two sub-genomes. Random chromosome means that the sequence was mapped to corresponding chromosome but the physical location was unknown. Length of chromosomes is also shown in the figure.

different members. The gene structure was displayed according to the five clades in the phylogenetic tree (Figure 4A). The gene length ranged from 153 to 4,195 base pairs (bp), and most genes were less than 1,000 bp. The exon number ranged from one to six, and 70.15% (47 out of 67) of the genes had only one exon (Figure 4B). Some genes displayed unique structures. For instance, only seven genes had untranslated region (UTR); *BnOFP15_3* and *BnOFP5_3* had upstream UTR, *BnOFP13_1*, *BnOFP16_1*, *BnOFP16_2*, and *BnOFP2_3* had downstream UTR; while *BnOFP19_2* had both upstream and downstream UTR (Figure 3B). This gene family had fewer introns, as well as great differences in intron length. *BnOFP17_1*, *BnOFP13_4*, *BnOFP13_3*, *BnOFP5_2*, and *BnOFP7_2* contained longer introns, and *BnOFP5_5* and *BnOFP3_2* had shorter introns. Some members derived from the same *Arabidopsis* gene showed the same gene length and structure, such as the *BnOFP8* and *BnOFP18* subfamily (Figure 4B).

We also analyzed the gene conserved domains, and identified a total of nine conserved domains. Ovate and ovate superfamily domain were detected in 68.66% of the *BnOFP* members (Figure 4C). Most members contained one domain, and a few members harbored two or three domains. Some domains were rarely detected, such as the P-loop_NTPase superfamily and PTZ00121 superfamily domains, which were only detected in *BnOFP5_2* and *BnOFP5_3*, respectively. However, 31.34% members had no ovate family domains, and seven members in clade I had no conserved domain. This loss of domain may be caused by genome polyploidization. These results may greatly help the research on the functional conservation and divergence of *BnOPFs* in *B. napus* evolution.

Differential expression of *BnOFPs* in different tissues

With the rapid development of RNA sequencing technology, transcriptome analysis has been widely used in gene functional research. The expression level of a gene is related to its functional performance. The FPKM (fragments per kilobase of exon model per million mapped fragments) value from RNA sequencing can effectively represent the gene expression levels and be used to analyze the biological processes in plant tissues in different stages and environments. In this study, a transcriptome analysis of twelve tissues in *B. napus* cultivar

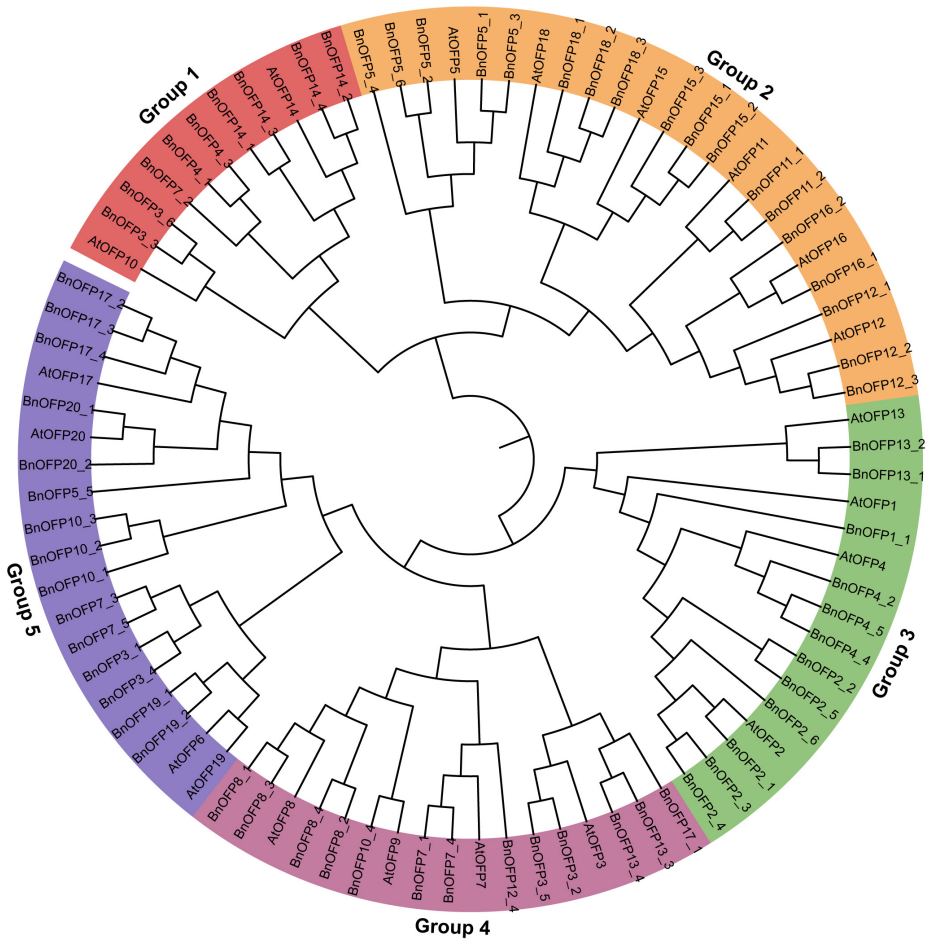


FIGURE 2
Phylogenetic relationship analysis of ovate family proteins (OFPs) in *Arabidopsis* and *Brassica napus*. The 20 *AtOFP* genes and 67 *BnOFP* genes were clustered into five groups. Members of each group were represented by different background color.

line 'ZS11' was performed to analyze the expression pattern of *BnOFPs*. We found that the expression of 67 *BnOFPs* varied greatly among the twelve tissues (Figure 5A and Supplementary Table 3). The majority of the genes had generally low expression (Supplementary Table 3). *BnOFP5_1*, *BnOFP5_6*, *BnOFP10_2*, *BnOFP10_4*, *BnOFP12_2*, and *BnOFP12_3* had relatively higher expression in the sepal, while *BnOFP14_2*, *BnOFP16_2*, *BnOFP7_1*, *BnOFP18_1*, *BnOFP14_3*, *BnOFP15_2*, and *BnOFP15_3* exhibited relatively higher expression in the bud. *BnOFPs* were lowly expressed in most tissues. For example, all the genes had low expression in the stem, while in the leaf and silique, only *BnOFP17_7* and *BnOFP2_1* had relatively high expression (Figure 5A). Ten genes showed no expression in any tissue, probably because they have lost their functions during evolution. A qRT-PCR experiment was then performed to verify the transcriptome data, and the results of five genes in four tissues are presented in Figure 5B. The expression of *BnOFP5_4*, *BnOFP5_6*, and *BnOFP16_1* was obviously higher in the sepal and lower in the new pistil, while that of *BnOFP19_2*, *BnOFP5_4*, and *BnOFP5_6* was obviously low in the bud and stem (Figure 5B). The qRT-PCR experimental results were consistent with the transcriptome results, verifying the reliability of the transcriptome data. The expression pattern of *BnOFPs* suggested the occurrence of functional divergence in this gene family of *B. napus* during evolution.

Association of *BnOFP13_2* with number of seeds per silique

Genome wide association study is a new approach to precisely predict the corresponding genes or QTLs involved in the regulation of complex traits of plants based on linkage disequilibrium (LD) (Nordborg and Weigel, 2008). In this study, GWAS for NSPS was performed with a natural population consisting of 204 accessions. The phenotype value ranged from 7.06 to 24.68, exhibiting large differences among different accessions (Figure 6B). Finally, a significant locus was identified

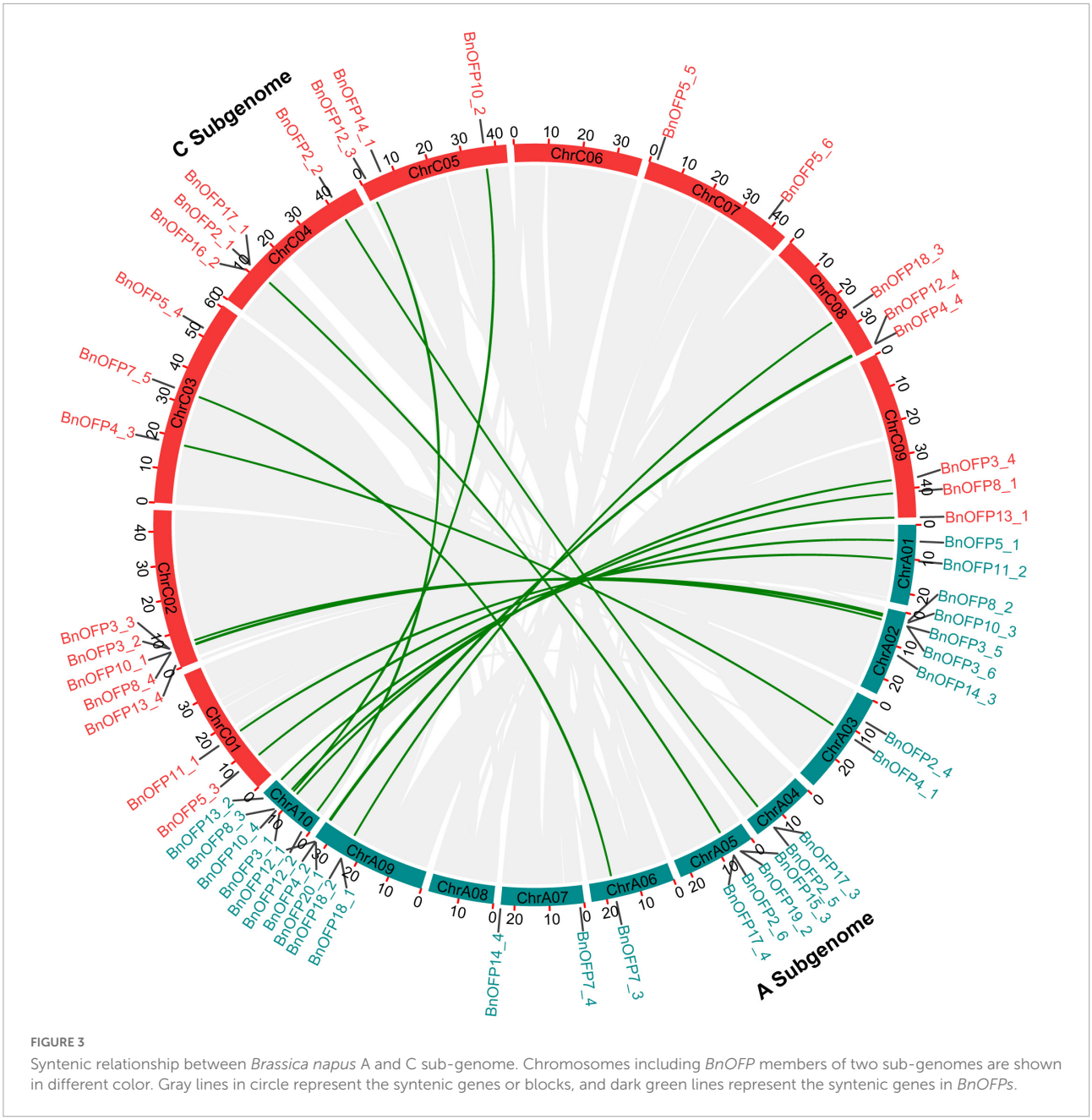


FIGURE 3
Syntenic relationship between *Brassica napus* A and C sub-genome. Chromosomes including *BnOFP* members of two sub-genomes are shown in different color. Gray lines in circle represent the syntenic genes or blocks, and dark green lines represent the syntenic genes in *BnOFPs*.

on chromosome A10. The 16.39–16.47 Mb block on A10 contained a *BnOFP* member, and the most significant SNP was located around *BnOFP13_2* (**Figure 6A** and **Supplementary Table 4**). Therefore, we speculated that *BnOFP13_2* might be associated with NSPS in *B. napus*, and then analyzed the SNP of *BnOFP13_2* in the 204 accessions. Two homozygous non-synonymous SNPs were obtained at the position of +40 and +604 bp of gene sequence. There were three classified haplotypes based on the two SNPs. Haplotype 1 comprised 14 accessions, and haplotype 2 and haplotype 3, respectively, had 141 and 23 accessions (**Figure 6C**). The accessions in haplotype 1, haplotype 2, and haplotype 3 had the average phenotype value of 16.50, 16.67, and 21.11, respectively, indicating that haplotype 3 had significantly higher NSPS than haplotype 1 and haplotype 2 (**Figure 6D**). These results suggested that *BnOFP13_2* is associated with NSPS.

Negative regulation of *BnOFP13_2* on number of seeds per silique in early silique development

To investigate how *BnOFP13_2* affects NSPS in *B. napus*, we detected the RNA profiles of *BnOFP13_2* at different silique

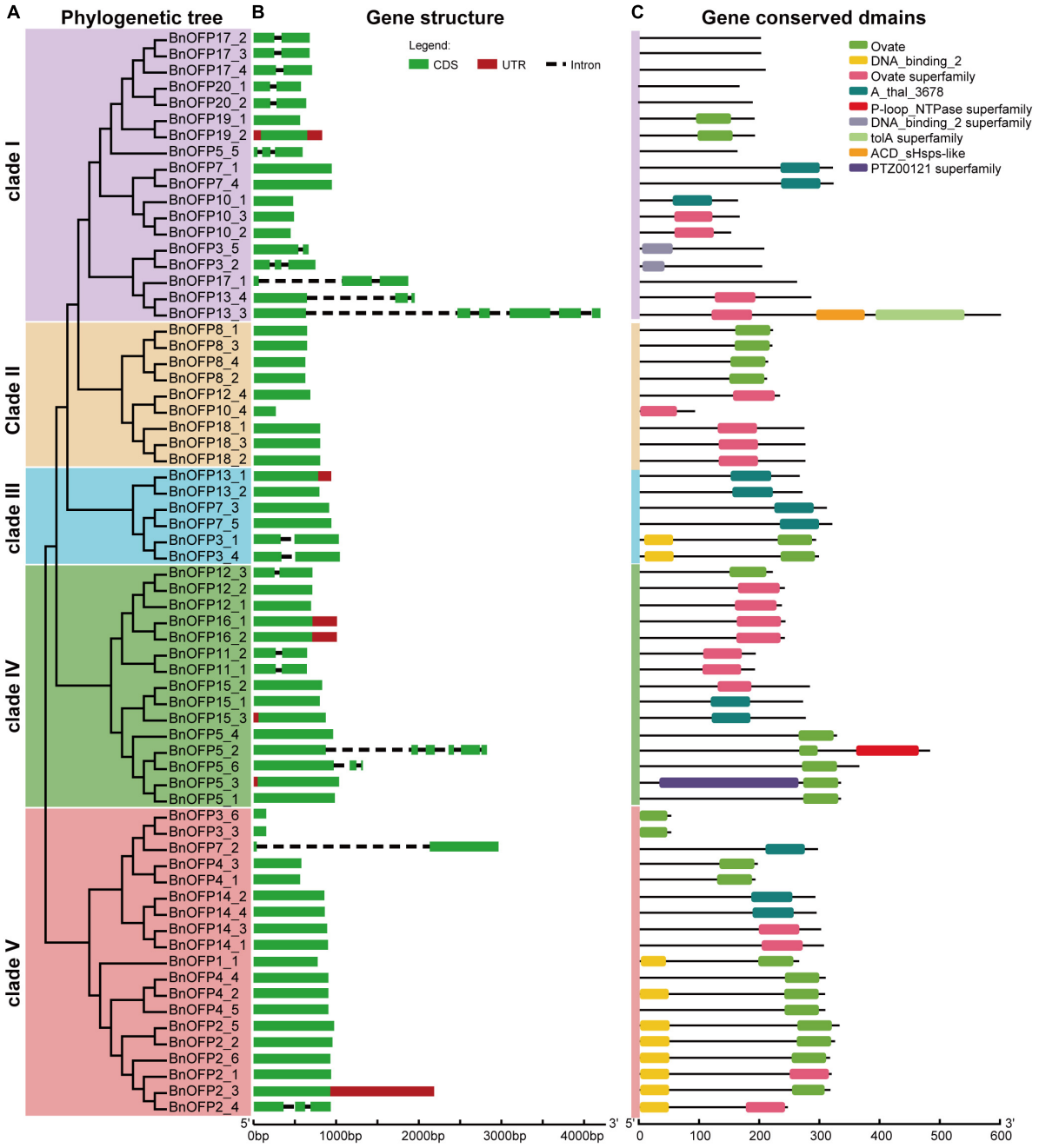


FIGURE 4
Gene structure and conserved domains of BnOFPs. **(A)** Phylogenetic tree of *BnOFP* genes. The 67 *BnOFP* genes were classified into five clades, each of which is illustrated with different colors. **(B)** Gene structure of the 67 *BnOFP* genes. Boxes represent the CDS and UTR section, and the lines indicate introns. **(C)** Gene conserved domains of the 67 *BnOFP* genes. Nine conserved domains are represented in different colored boxes, and the line means non-conserved sequence. Length scale is presented at the bottom.

developmental stages in two varieties, respectively, with high NSPS ('ZC520') and low NSPS ('ZC519'). 'ZC520' averagely had 24 seeds per silique and 'ZC519' only had 18 seeds per silique (Figures 7A,B). The siliques at 7 days after pollination (7 DAP), 14, 21, 28, and 35 DAP were collected to analyze the dynamics of the expression level of *BnOFP13_2* (Figure 7C). At 7 and 14 DAP, the expression of *BnOFP13_2* was significantly higher in 'ZC520' than in 'ZC519' (Figure 7D). From 21 to 35 DAP, the expression level of *BnOFP13_2* was constantly high without significant difference between the two varieties (Figure 7D). In 'ZC520' (high-NSPS variety), the expression of *BnOFP13_2* was low in the early developmental stage

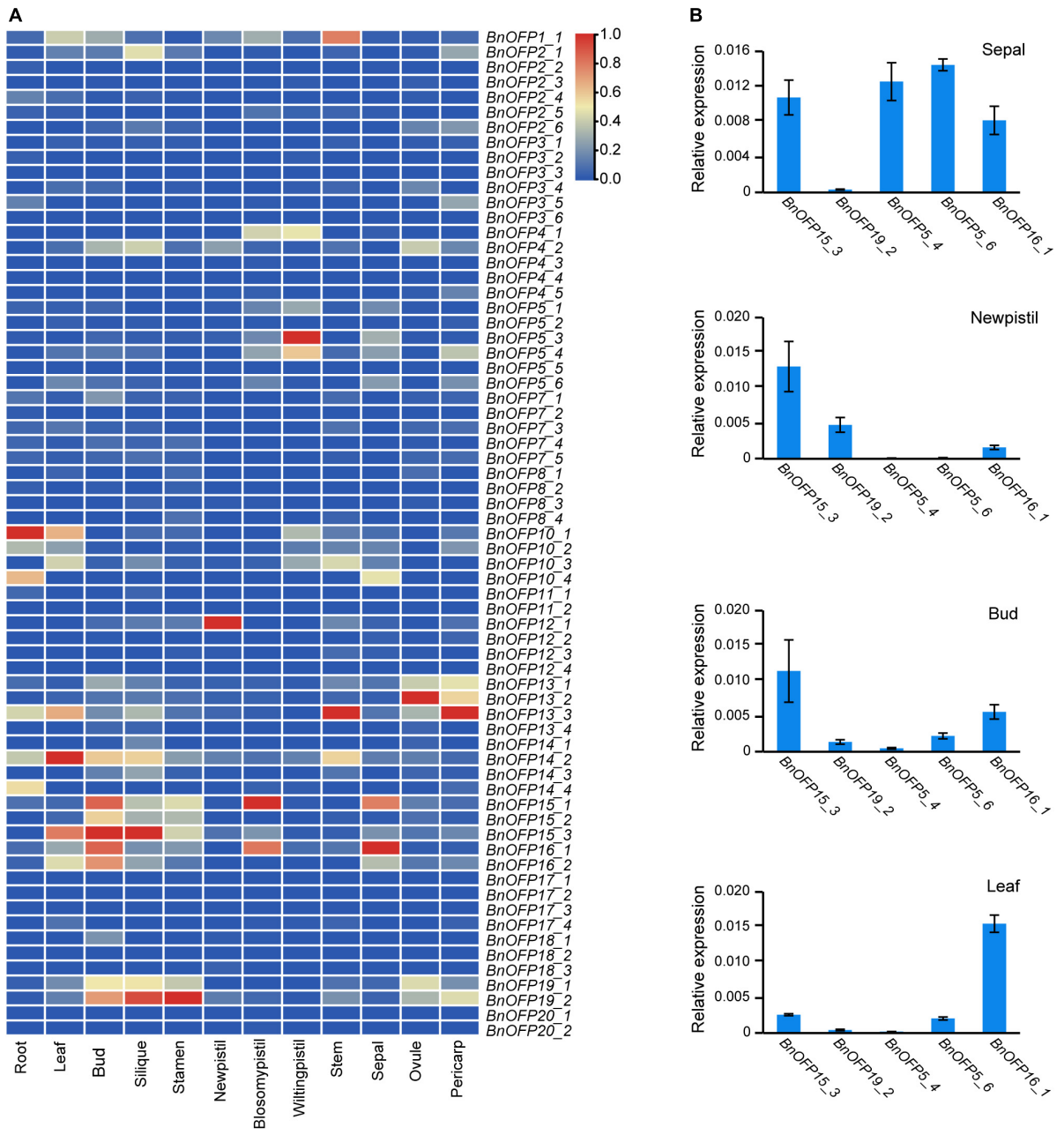


FIGURE 5
Expression patterns of *BnOFP* genes. **(A)** Heatmap of gene expression patterns. Expression levels of 67 *BnOFP* genes in twelve tissues of 'ZS11' obtained from FPKM of transcriptome analysis. **(B)** Verification of the expression level by qRT-PCR. Five genes and four tissues were used to verify the expression level in transcriptome analysis by qRT-PCR.

of silique, which then increased slowly until 21 DAP and remained constant until maturity; while in 'ZC519' (low-NSPS variety), *BnOFP13_2* showed a high expression level at almost all developmental stages (Figure 7D). These results suggested that the accumulation of *BnOFP13_2* transcripts at the early developmental stage of silique is negatively correlated with NSPS. Therefore, we speculated that *BnOFP13_2* suppresses the formation or development of ovule, which is similar to the previous finding in *Arabidopsis*, but the specific molecular mechanism remains to be further studied.

Discussion

More than 20 years ago, OVATE was a well-known QTL for its effect in controlling the pear shape of tomato fruit,

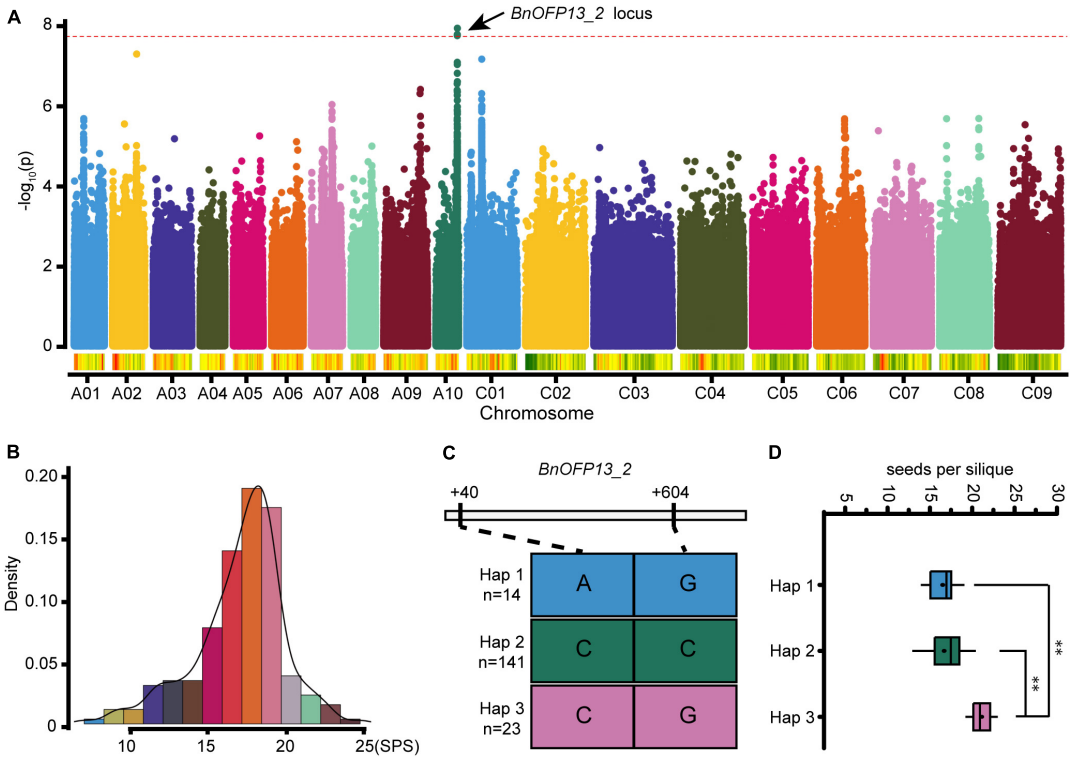


FIGURE 6
Genome wide association study (GWAS) and haplotype analysis. **(A)** Manhattan map of GWAS. The red dotted line represents the threshold value 7.72. Dots above the red dotted line indicate the significant loci associated with NSPS. **(B)** Histogram of frequency distribution. Statistical analysis of NSPS phenotype used in GWAS. Phenotype variations of 204 accessions are presented. **(C)** Haplotype analysis of *BnOFP13_2*. All the accessions were classified into three haplotypes based on the two non-synonymous SNPs of *BnOFP13_2*. **(D)** Haplotype analysis. Statistical analysis of accessions in each haplotype, and the phenotype value shows significant differences. $^{**}p < 0.01$ in ANOVA test.

which was then mapped and cloned on chromosome 2 (Ku et al., 2001; Liu et al., 2002). The protein and RNA profiles have suggested that OVATE is a novel class of proteins named as OFPs (ovate family proteins) (Liu et al., 2002). Owing to their effects on fruits, *OFPs* have been studied in many fruit crops, and the results have indicated that *OFPs* have great influence on a variety of aspects in plants, including fruit shape (Liu et al., 2002; Tsaballa et al., 2011; Cirilli and Rossini, 2021; Han et al., 2022), tuberous root shape (Wang et al., 2020), ovule development (Pagnussat et al., 2007), secondary cell wall formation (Li et al., 2011; Wang et al., 2011), vasculature development (Schmitz et al., 2015), fruit ripening (Liu et al., 2015b), DNA repair (Wang et al., 2010), floral organs, compound leaf and silique (Liu et al., 2002; Wang et al., 2011). Here, we briefly summarized the functions of OFPs in *Arabidopsis* and some common crops (**Figure 8A**), and we believe that more functions of OFPs will be gradually detected in plants. There has been no report about the functions of *OFPs* in *B. napus*. In this study, 67 *BnOFP* members were checked with gene ontology (GO) annotations (**Figure 8B**). As a result, 63 members were involved in biological regulation, cellular process, metabolic process, negative regulation of biological process and regulation of biological process; 52 members were related to cell part and organelle component; and 12 members had the binding molecular function (**Figure 8B**). These results suggested that *BnOFPs* play different roles in plant development. We further identified and characterized the *OFPs* in *B. napus*, as well as analyzed their phylogenetic relationship with the orthologs in *Arabidopsis* and the syntenic relationship between the two sub-genomes. We also analyzed the gene structure, conserved domain and expression pattern, and the differences in the gene structure and expression could provide a theoretical basis for functional research of this gene family. In addition, a novel function of *OFPs* was identified through GWAS, and haplotype analysis suggested that *BnOFP13_2* is significantly associated with NSPS. Further expression profiling at the silique developmental stage indicated that *BnOFP13_2* negatively regulates NSPS during early silique development. Our results may provide important information for rapeseed breeding.

In *Arabidopsis*, *AtOFP1* was firstly identified as a transcriptional repressor of cell elongation in many organs. An increase in the expression of *AtOFP1* reduced the length of hypocotyl, inflorescence stem, cauline leaf, rosette leaf,

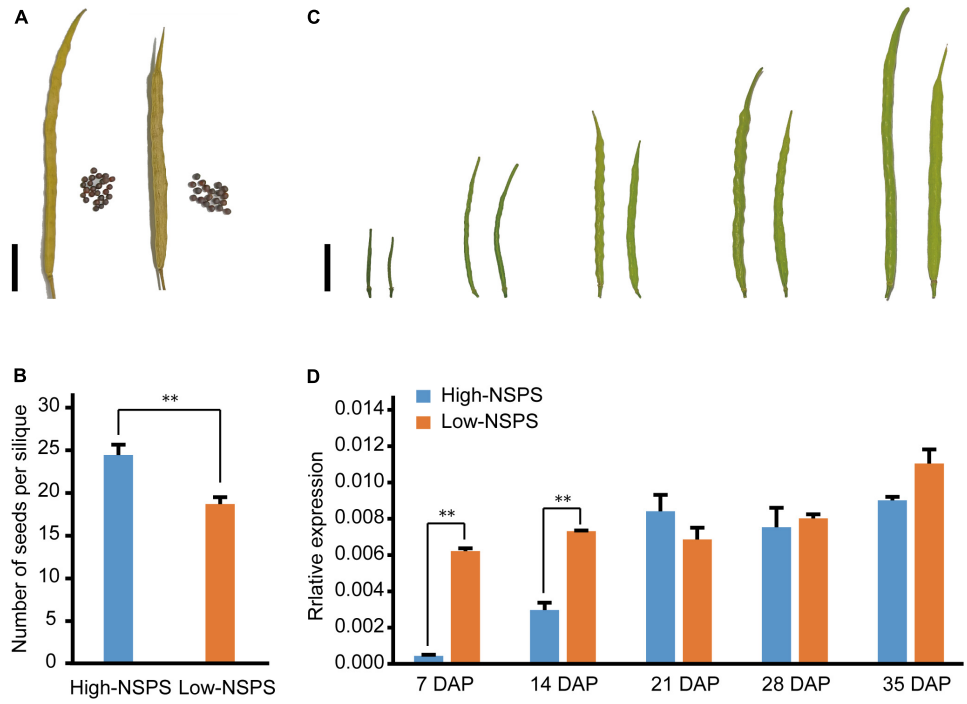


FIGURE 7
Phenotype of siliques at different developmental stages and expression levels of BnOFP13_2 in high-NSPS variety and low-NSPS variety. **(A)** Mature silique of high-NSPS variety (left) and low-NSPS variety (right), bar: 2 cm. **(B)** Statistical analysis of NSPS in high-NSPS variety and low-NSPS variety. **(C)** Phenotype of high-NSPS variety and low-NSPS variety siliques at 7, 14, 21, 28, and 35 DAP, bar: 2 cm. **(D)** Relative expression levels of *BnOFP13_2* in high-NSPS variety and low-NSPS variety silique at different developmental stages. $^{**}p < 0.01$.

floral organ, and silique (Wang et al., 2007). Overexpression of *AtOFP2*, *AtOFP4*, and *AtOFP7* led to similar phenotypes to overexpression of *AtOFP1* such as round and curled leaves, implying that these three genes are also transcriptional repressors of cell elongation (Wang et al., 2007; Li et al., 2011). Overexpression of *AtOFP6* and *AtOFP8* resulted in thick and cyan rosette leaves in plants, while that of *AtOFP13*, *AtOFP15*, and *AtOFP16* led to short and blunt-ended siliques (Wang et al., 2011). In this study, high expression of *BnOFP13_2* was detected at early developmental stage of siliques in low-NSPS variety. Therefore, it can be speculated that *BnOFP13_2* negatively regulates NSPS by reducing silique length and inhibiting ovule development at early developmental stage.

Since *OFPs* are negative regulators of plant growth and development, the underlying mechanisms have received great attention. Although the mechanisms remain largely unknown, there has been some evidence implying that *OFPs* may directly regulate the expression of the target genes. In *Arabidopsis*, AtOFP1 contains a putative NLS domain and is located in the nucleus, and a yeast one-hybrid experiment indicated that *AtOFP1* represses the expression of the reporter gene. These results suggest that *AtOFP1* is a transcription repressor directly regulating gene expression (Wang et al., 2007, 2011). In addition, *AtOFP1* also contains the LxLxL domain, which was also found in ERF transcription factors and is necessary for transcriptional repression (Hiratsu et al., 2003). On the other hand, *OFPs* were found to form a complex to perform their functions *via* interacting with other proteins. For instance, AtOFP4 and AtOFP1 interact with KNAT7 and BLH6 to regulate secondary cell wall formation (Li et al., 2011); AtOFP1 coordinates with ATH1 to affect flowering time and stem growth (Zhang et al., 2018); AtOFP1 has interaction with BLH3 to regulate the transition from vegetative phase to reproductive phase (Zhang et al., 2016); AtOFP5 cooperates with KANT2 and BLH1 in regulating female gametophyte development (Pagnussat et al., 2007); MaOFP1 interacts with MuMADS1 to regulate fruit ripening (Liu et al., 2015b); GhOFP4 coordinates with GhKNL1 in mediating secondary cell wall formation (Gong et al., 2014); and OsOFP2 interacts with OsKNAT7 and BLH6-like1 to mediate vascular development (Schmitz et al., 2015). With a better understanding of these negative regulators, *OFPs* can be better used for crop improvement *via* the gene editing technology.

In plants, polyploidization occurs at high frequency. It has been demonstrated that about 25–30% of the existing flowering plants are polyploids and have not been diploidized yet (Wood et al., 2009; Scarpino et al., 2014). Polyploids are generally classified into autopolyploids and allopolyploids, with

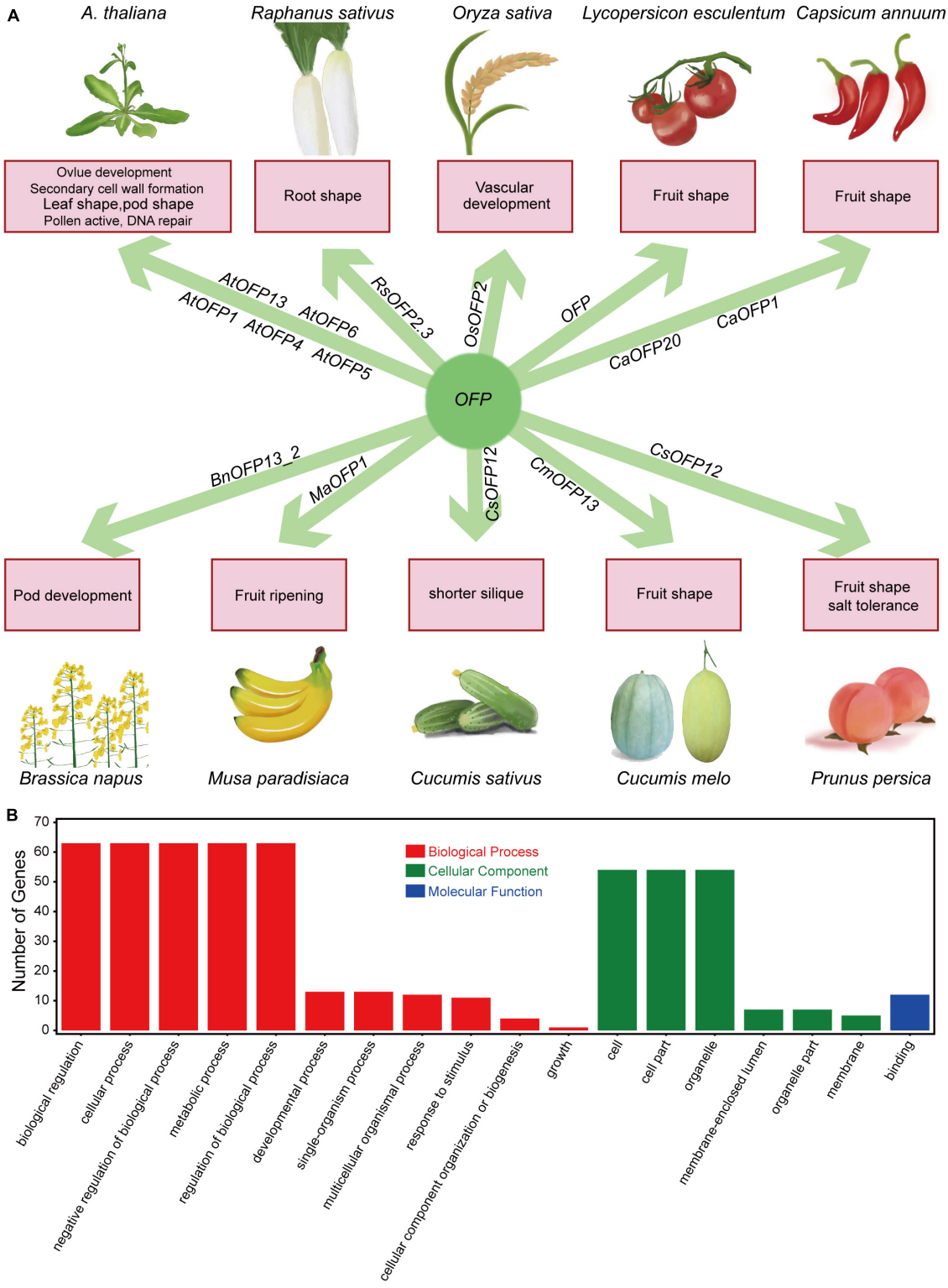


FIGURE 8
Brief summary of *OFP* functions and GO annotations of the 67 *BnOFP* genes. **(A)** Function summary of *OFPs* in common crops. The crop pictures are painted. **(B)** GO annotations of *OFP* genes in *B. napus*. Gene numbers and annotations in biological processes, cellular components and molecular functions are shown in the figure.

the former being formed through doubling of one chromosome set, while the latter being formed *via* hybridization or merging of different chromosome sets in different species (Ramsey and Schemske, 1998; Barker et al., 2016; Van de Peer et al., 2017). Allopolyploidy is involved in epigenetic remodeling and changes in gene expression, which will contribute to a higher adaptive potential (Yoo et al., 2014; Hu et al., 2015b; Van de Peer et al., 2017). It is widely accepted that polyploids have increased the mutational robustness and adaptability compared with diploids. Polyploids can overcome sympatrically speciate and cytotype minority exclusion for rapid adaptation, which is also owing to their higher genomic plasticity than diploids (te Beest et al., 2012; Schoenfelder and Fox, 2015). Gene duplication generally has four modes, including the tetraploid, segmental, tandem, and transpositional modes, with each mode retaining genes in a biased method (Neufeld et al., 1991; Freeling, 2009). After genome or segment duplication, some duplicated genes are often lost (Freeling, 2009). Polyploidization or WGD is an important driving force for evolution in animals and plants (Abdel-haleem, 2007; Edger and Pires, 2009), and the success of angiosperms is partially attributed to WGD (De Bodt et al., 2005). In *Arabidopsis*, *OFPs* are involved in different functions. After polyploidization into *B. napus*, the *OFP* members have experienced gene doubling, loss and functional divergence, which may lead to the appearance of novel gene structure, conserved domains and functions. In this study, seven *BnOFP* members were detected to have no ovate or related domains (**Figure 4C**), which might have been caused by loss-of-function during polyploidization.

Conclusion

In this study, we identified 67 *BnOFP* genes at the genome level, analyzed their evolution from *Arabidopsis* to *B. napus* and compared the gene structure and conserved domains between copies. We also identified a new potential locus significantly correlated with NSPS, which is a novel function of *OFPs*. RNA profiling in different stages of silique development suggested that *BnOFP13_2* negatively regulates the NSPS by decreasing the silique length and inhibiting the ovule development at early developmental stage. Our findings provide evidence for the functional diversity of *OFP* gene family and new implications for oilseed rape breeding.

Materials and methods

Plant materials and phenotype identification

The 204 natural accessions and 'ZS11' used for GWAS analysis and qRT-PCR experiment were obtained from the Key Laboratory of Biology and Genetic Improvement of Oil Crops at OCRI and planted in our experimental field in WuHan, HuBei province, China. These materials used for GWAS were collected from all over the world, including 52 spring accessions, 40 winter accessions and 112 semi-winter accessions, and were cultivated under natural growing conditions. The phenotype data were collected in 1 year (2017–2018). Planting was conducted with twelve plants in a row and a row spacing of 30 cm. All varieties were self-bred for many generations and were pure inbred lines. Ten siliques from different parts of the main inflorescence were collected to evaluate the NSPS. The phenotype of each variety was represented by the average of three biological replicates.

Identification of *BnOFPs* in *Brassica napus*

The *OFP* gene protein sequences of *Arabidopsis* were obtained from the TAIR database[1], and used to search for the *OFP* genes in *B. napus*. The annotation and genome information of *B. napus* cultivar 'Darmor-bzh' and corresponding orthologous genes in *Arabidopsis* were obtained from the BRAD (Brassicaceae Database) database[2] (Chalhoub et al., 2014). The HMMER3.0[3] was used to search for OFP genes in *B. napus* (E value was set to 1e-5).

Gene structure and conserved domain analysis

Gene structure files of *BnOFPs* were downloaded from EnsemblPlants database[4], and illustrated through $GSDS_{2.0}$ (Gene Structure Display Server) online tools[5] (Hu et al., 2015a). Gene conserved domains were analyzed through CD-search in NCBI (National Center for Biotechnology Information) database[6] using the amino acid sequence, and illustrated through TBtools (Chen et al., 2020).

Phylogenetic and syntenic relationship analysis

The protein sequences of *OFPs* in *Arabidopsis* and *B. napus* were used to construct a phylogenetic tree using the MEGA

1 https://www.arabidopsis.org/
2 http://brassicadb.cn/#/
3 http://www.hmmer.org/
4 http://plants.ensembl.org/index.html
5 http://gsds.gao-lab.org/
6 https://www.ncbi.nlm.nih.gov/

software. The beautification of phylogenetic tree was conducted with iTOL online tools.[7] The syntenic relationship of gene or block between two sub-genomes were obtained from the BRAD database (Chalhoub et al., 2014). The syntenic relationship figure was drawn through TBtools (Chen et al., 2020).

Transcriptome analysis and qRT-PCR analysis

The transcriptome data of twelve tissues in 'ZS11' used for the expression pattern analysis were obtained from our lab and have been already published (SRA accession: PRJNA474576) (Li et al., 2019b). The heatmap was drawn through TBtools (Chen et al., 2020). The sepal, leaf, bud, and newpistil samples used for RNA extraction were collected in experimental field. The silique samples were collected at 7, 14, 21, 28, and 35 DAP. All the samples had three biological replicates. The TRIzol reagent (Invitrogen, Carlsbad, CA, United States) was used for total RNA extraction. About 2 μg RNA was used to reverse transcribe using the PrimeScript^{TM} RT reagent Kit (TaKaRa Co., Ltd., Beijing, China). The *B. napus* gene *BnaA10g22340D* was used as the reference gene. The relative expression was evaluated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Genome wide association study analysis in natural population

The genomic DNA of all plants was extracted from tender rosette leaves using the modified CTAB method (Allen et al., 2006). The SNP data were obtained from 7 × re-sequencing data mapped to the reference genome 'Darmor-bzh'. Re-sequencing was performed by the commercial Illumina HiSeq XTen service (BGI-Shenzhen, China). The SNPs were finalized under the minor allele frequency <0.05, and 2611513 valid SNPs were obtained at last, the original SNPs were obtained from published data of our lab (Ding et al., 2020). GWAS for NSPS was performed with the R package using the general linear model (Yin et al., 2021). The threshold value was set to $p < -\log_{10}(0.05/N)$, where N represents the number of used SNPs.

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**.

7 https://itol.embl.de/

Author contributions

JL, XhC, and JH designed the research. SL supervised the research. JL and YW performed the experiments and wrote the manuscript. YW, XbC, and XZ collected the data. MX helped analyze the data. LL and YL provided the plant materials. XhC and JH revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Key Research and Development Program of China (2021YFD1600500), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ZDRW202105 and CAAS-ASTIP-2013-OCRI), Central Public-interest Scientific Institution Basal Research Fund (Y2020YJ03 and 1610172021001), and China Agriculture Research System of MOF and MARA (CARS-12).

Acknowledgments

Thanks to all the laboratory members for investigating the phenotype in GWAS analysis, thanks to Minqiang Tang for providing the SNP data, thanks to XhC and YL for organizing the phenotype investigation, thanks to Luyao Fan for painting the crop pictures.

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.2022.962592/full#supplementary-material

References

Abdel-haleem, H. (2007). The origins of genome architecture. *J. Hered.* 98, 633–634. doi: 10.1093/jhered/esm073

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384

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

Barker, M. S., Arrigo, N., Baniaga, A. E., Zheng, L., and Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698

Borovsky, Y., Raz, A., Doron-Faigenboim, A., Zemach, H., Karavani, E., and Paran, I. (2022). Pepper fruit elongation is controlled by capsicum annuum ovate family protein 20. *Front. Plant Sci.* 12:815589. doi: 10.3389/fpls.2021.815589

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

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

Cirilli, M., and Rossini, L. (2021). Many candidates for a single chair: A critical review of the genetic determinant of flat fruit shape trait in peach (*Prunus persica* L. Batsch). *Tree Genet. Genom.* 17:34. doi: 10.1007/s11295-021-01515-w

Clarke, J. M., and Simpson, G. M. (1978). Influence of irrigation and seeding rates on yield and yield components of *Brassica napus* cv. tower. *Can. J. Plant Sci.* 58, 731–737. doi: 10.4141/cjps78-108

D'Avino, L., Dainelli, R., Lazzeri, L., and Spugnoli, P. (2015). The role of co-products in biorefinery sustainability: Energy allocation versus substitution method in rapeseed and carinata biodiesel chains. *J. Clean. Prod.* 94, 108–115. doi: 10.1016/j.jclepro.2015.01.088

De Bodt, S., Maere, S., and Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* 20, 591–597. doi: 10.1016/j.tree.2005.07.008

Ding, L.-N., Li, M., Guo, X.-J., Tang, M.-Q., Cao, J., Wang, Z., et al. (2020). Arabidopsis GDSL1 overexpression enhances rapeseed *Sclerotinia sclerotiorum* resistance and the functional identification of its homolog in *Brassica napus*. *Plant Biotechnol. J.* 18, 1255–1270. doi: 10.1111/pbi.13289

Edger, P. P., and Pires, J. C. (2009). Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes. *Chromosom. Res.* 17:699. doi: 10.1007/s10577-009-9055-9

Felten, D., Fröba, N., Fries, J., and Emmerling, C. (2013). Energy balances and greenhouse gas-mitigation potentials of bioenergy cropping systems (*Miscanthus*, rapeseed, and maize) based on farming conditions in Western Germany. *Renew. Energy* 55, 160–174. doi: 10.1016/j.renene.2012.12.004

Ferrándiz, C., Pelaz, S., and Yanofsky, M. F. (1999). Control of carpel and fruit development in *Arabidopsis*. *Annu. Rev. Biochem.* 68, 321–354. doi: 10.1146/annurev.biochem.68.1.321

Freeling, M. (2009). Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Fu, Y., Wei, D., Dong, H., He, Y., Cui, Y., Mei, J., et al. (2015). Comparative quantitative trait loci for silique length and seed weight in *Brassica napus*. *Sci. Rep.* 5:14407. doi: 10.1038/srep14407

Gong, S. Y., Huang, G. Q., Sun, X., Qin, L. X., Li, Y., Zhou, L., et al. (2014). Cotton KNL1, encoding a class II KNOX transcription factor, is involved in regulation of fibre development. *J. Exp. Bot.* 65, 4133–4147. doi: 10.1093/jxb/eru182

Hackbusch, J., Richter, K., Müller, J., Salamini, F., and Uhrig, J. F. (2005). A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4908–4912. doi: 10.1073/pnas.0501181102

Han, L.-j, Song, X.-f, Wang, Z.-y, Liu, X.-f, Yan, L.-y, Han, D.-g, et al. (2022). enome-wide analysis of OVATE family proteins in cucumber (*Cucumis sativus* L.). *J. Integr. Agric.* 21, 1321–1331. doi: 10.1016/s2095-3119(21)63788-7

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J.* 34, 733–739. doi: 10.1046/j.1365-313x.2003.01759.x

Hopping, M. E., Monselise, S. P., Monselise, S., and Hopping, M. E. (1986). *CRC handbook of fruit set and development*. Boca Raton, FL: CRC Press Inc. doi: 10.1201/9781351073042

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

Hu, G., Koh, J., Yoo, M. J., Chen, S., and Wendel, J. F. (2015b). Gene-expression novelty in allopolyploid cotton: A proteomic perspective. *Genetics* 200, 91–104. doi: 10.1534/genetics.115.174367

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Jiao, Y., Zhang, K., Cai, G., Yu, K., Amoo, O., Han, S., et al. (2021). Fine mapping and candidate gene analysis of a major locus controlling ovule abortion and seed number per silique in *Brassica napus* L. *Theor. Appl. Genet.* 134, 2517–2530. doi: 10.1007/s00122-021-03839-6

Ku, H. M., Liu, J., Doganlar, S., and Tanksley, S. D. (2001). Exploitation of *Arabidopsis*-tomato synteny to construct a high-resolution map of the ovatecontaining region in tomato chromosome 2. *Genome* 44, 470–475. doi: 10.1139/g01-024

Li, E., Wang, S., Liu, Y., Chen, J. G., and Douglas, C. J. (2011). OVATE FAMILY PROTEIN4 (OFP4) interaction with KNAT7 regulates secondary cell wall formation in *Arabidopsis thaliana*. *Plant J.* 67, 328–341. doi: 10.1111/j.1365-313X.2011.04595.x

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., Li, J., Song, J., Zhao, B., Guo, C., Wang, B., et al. (2019a). An auxin signaling gene BnaA3.IAA7 contributes to improved plant architecture and yield heterosis in rapeseed. *New Phytol.* 222, 837–851. doi: 10.1111/nph.15632

Li, Y., Dong, C., Hu, M., Bai, Z., Tong, C., Zuo, R., et al. (2019b). Identification of Flower-Specific Promoters through Comparative Transcriptome Analysis in Brassica napus. *Intl. J. Mol. Sci.* 20:5949. doi: 10.3390/ijms20235949

Liu, D., Sun, W., Yuan, Y., Zhang, N., Hayward, A., Liu, Y., et al. (2014). Phylogenetic analyses provide the first insights into the evolution of OVATE family proteins in land plants. *Ann. Bot.* 113, 1219–1233. doi: 10.1093/aob/mcu061

Liu, J., Van Eck, J., Cong, B., and Tanksley, S. D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13302–13306. doi: 10.1073/pnas.162485999

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

Liu, J., Zhang, J., Hu, W., Miao, H., Zhang, J., Jia, C., et al. (2015b). Banana Ovate family protein MaOFP1 and MADS-box protein MuMADS1 antagonistically regulated banana fruit ripening. *PLoS One* 10:e0123870. doi: 10.1371/journal.pone.0123870

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

Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., et al. (2017). Incorporating pleiotropic quantitative trait loci in dissection of complex traits: Seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585. doi: 10.1007/s00122-017-2911-7

Ma, J., Li, C., Zong, M., Qiu, Y., Liu, Y., Huang, Y., et al. (2021). CmFSI8/CmOFP13 encoding an OVATE family protein controls fruit shape in melon. *J. Exp. Bot.* 73, 1370–1384. doi: 10.1093/jxb/erab510

Monforte, A. J., Diaz, A., Cano-Delgado, A., and van der Knaap, E. (2014). The genetic basis of fruit morphology in horticultural crops: Lessons from tomato and melon. *J. Exp. Bot.* 65, 4625–4637. doi: 10.1093/jxb/eru017

Nagaharu, U. (1935). Genome analysis in *Brassica carinata* with special reference to the experimental formation of *Brassica napus*, a peculiar mode of fertilization. *Japanese J. Bot.* 7, 389–452.

Neufeld, T. P., Carthew, R. W., and Rubin, G. M. (1991). Evolution of gene position: Chromosomal arrangement and sequence comparison of the Drosophila melanogaster and Drosophila virilis sina and Rh4 genes. *Proc. Natl. Acad. Sci. U.S.A.* 88, 10203–10207. doi: 10.1073/pnas.88.22.10203

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. *Nature* 456, 720–723. doi: 10.1038/nature07629

Pagnussat, G. C., Yu, H. J., and Sundaresan, V. (2007). Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. *Plant Cell* 19, 3578–3592. doi: 10.1105/tpc.107.054890

Price, H. L., and Drinkard, A. W. (1908). Inheritance in tomato hybrids. *Zeitschrift für induktive Abstammungs- und Vererbungslehre* 1, 402–403. doi: 10.1007/BF01990617

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Reinhardt, D., and Kuhlemeier, C. (2002). Plant architecture. *EMBO Rep.* 3, 846–851. doi: 10.1093/embo-reports/kvf177

Scarpino, S. V., Levin, D. A., and Meyers, L. A. (2014). Polyploid formation shapes flowering plant diversity. *Am. Nat.* 184, 456–465. doi: 10.1086/677752

Schmitz, A. J., Begcy, K., Sarath, G., and Walia, H. (2015). Rice Ovate Family Protein 2 (OFP2) alters hormonal homeostasis and vasculature development. *Plant Sci.* 241, 177–188. doi: 10.1016/j.plantsci.2015.10.011

Schoenfelder, K. P., and Fox, D. T. (2015). The expanding implications of polyploidy. *J. Cell Biol.* 209, 485–491. doi: 10.1083/jcb.201502016

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

Tan, Q., Jiang, S., Wang, N., Liu, X., Zhang, X., Wen, B., et al. (2021). OVATE Family Protein PpOFP1 Physically Interacts With PpZFHD1 and Confers Salt Tolerance to Tomato and Yeast. *Front. Plant Sci.* 12:759955. doi: 10.3389/fpls.2021.759955

te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubesová, M., et al. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Anna. Bot.* 109, 19–45. doi: 10.1093/aob/mcr277

Tsaballa, A., Pasentsis, K., Darzentas, N., and Tsaftaris, A. S. (2011). Multiple evidence for the role of an Ovate-like gene in determining fruit shape in pepper. *BMC Plant Biol.* 11:46. doi: 10.1186/1471-2229-11-46

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

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

Wang, G., Zhang, X., Huang, W., Xu, P., Lv, Z., Zhao, L., et al. (2021). Increased seed number per silique in *Brassica juncea* by deleting cis-regulatory region affecting BjCLV1 expression in carpel margin meristem. *Plant Biotechnol. J.* 19, 2333–2348. doi: 10.1111/pbi.13664

Wang, S., Chang, Y., Guo, J., and Chen, J. G. (2007). arabidopsis ovate family protein 1 is a transcriptional repressor that suppresses cell elongation. *Plant J.* 50, 858–872. doi: 10.1111/j.1365-313X.2007.03096.x

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B. E., and Chen, J. G. (2011). Arabidopsis ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. *PLoS One* 6:e23896. doi: 10.1371/journal.pone.0023896

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. doi: 10.1186/s12870-016-0759-7

Wang, Y. K., Chang, W. C., Liu, P. F., Hsiao, M. K., Lin, C. T., Lin, S. M., et al. (2010). Ovate family protein 1 as a plant Ku70 interacting protein involving in DNA double-strand break repair. *Plant. Mol. Biol.* 74, 453–466. doi: 10.1007/s11103-010-9685-5

Wang, Y., and Li, J. (2008). Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* 59, 253–279. doi: 10.1146/annurev.arplant.59.032607.092902

Wang, Y., Wang, Q., Hao, W., Sun, H., and Zhang, L. (2020). Characterization of the OFP gene family and its putative involvement of tuberous root shape in radish. *Int. J. Mol. Sci.* 21:1293. doi: 10.3390/ijms21041293

Williams, W. (1965). Evolution of Crop Plants. *Nature* 206, 63–64. doi: 10.1038/206063b0

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106

Yang, Y., Shen, Y., Li, S., Ge, X., and Li, Z. (2017). High density linkage map construction and QTL detection for three silique-related traits in orychophragmus violaceus derived brassica napus population. *Front. Plant Sci.* 8:1512. doi: 10.3389/fpls.2017.01512

Yang, Y., Shi, J., Wang, X., Liu, G., and Wang, H. (2016). Genetic architecture and mechanism of seed number per pod in rapeseed: Elucidated through linkage and near-isogenic line analysis. *Sci. Rep.* 6:24124. doi: 10.1038/srep24124

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A Memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteom. Bioinform.* 19, 619–628. doi: 10.1016/j.gpb.2020.10.007

Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Zhang, L., Sun, L., Zhang, X., Zhang, S., Xie, D., Liang, C., et al. (2018). OFP1 Interaction with ATH1 regulates stem growth, flowering time and flower basal boundary formation in *Arabidopsis*. *Genes (Basel)* 9, 399. doi: 10.3390/genes9080399

Zhang, L., Zhang, X., Ju, H., Chen, J., Wang, S., Wang, H., et al. (2016). Ovate family protein1 interaction with BLH3 regulates transition timing from vegetative to reproductive phase in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 470, 492–497. doi: 10.1016/j.bbrc.2016.01.135

Zhu, Y., Ye, J., Zhan, J., Zheng, X., Zhang, J., Shi, J., et al. (2020). Validation and characterization of a seed number per silique quantitative trait locus qSN.A7 in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 11:68. doi: 10.3389/fpls.2020.00068

Zu, S. H., Jiang, Y. T., Chang, J. H., Zhang, Y. J., Xue, H. W., and Lin, W. H. (2022). Interaction of brassinosteroid and cytokinin promotes ovule initiation and increases seed number per silique in *Arabidopsis*. *J. Integr. Plant Biol.* 64, 702–716. doi: 10.1111/jipb.13197

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fpls.2022.944763

OPEN ACCESS

EDITED BY
Ryo Fujimoto,
Kobe University, Japan

REVIEWED BY
Intikhab Alam,
Fujian Agriculture and Forestry University, China
Lisong Ma,
Hebei Agricultural University, China
Tongkun Liu,
Nanjing Agricultural University, China

*CORRESPONDENCE
Shitou Xia
xstone0505@hunau.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 15 May 2022
ACCEPTED 11 July 2022
PUBLISHED 16 August 2022

CITATION
Li W, Lu J, Yang C and Xia S (2022) Identification of receptor-like proteins induced by *Sclerotinia sclerotiorum* in *Brassica napus*. *Front. Plant Sci.* 13:944763. doi: 10.3389/fpls.2022.944763

Identification of receptor-like proteins induced by *Sclerotinia sclerotiorum* in *Brassica napus*

Wei Li[1,2†], Junxing Lu[2†], Chenghuizi Yang[1] and Shitou Xia[1*]

[1]Hunan Provincial Key Laboratory of Phytohormones and Growth Development, College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha, China, [2]College of Life Science, Chongqing Normal University, Chongqing, China

Heightening the resistance of plants to microbial infection is a widely concerned issue, especially for economical crops. Receptor-like proteins (RLPs), typically with tandem leucine-rich repeats (LRRs) domain, play a crucial role in mediating immune activation, being an indispensable constituent in the first layer of defense. Based on an analysis of orthologs among *Brassica rapa*, *Brassica oleracea,* and *Brassica napus* using *Arabidopsis thaliana* RLPs as a reference framework, we found that compared to *A. thaliana*, there were some obvious evolutionary diversities of RLPs among the three *Brassicaceae* species. *BnRLP* encoding genes were unevenly distributed on chromosomes, mainly on chrA01, chrA04, chrC03, chrC04, and chrC06. The orthologs of five *At*RLPs (*At*RLP3, *At*RLP10, *At*RLP17, *At*RLP44, and *At*RLP51) were highly conserved, but retrenchment and functional centralization occurred in *Brassicaceae* RLPs during evolution. The RLP proteins were clustered into 13 subgroups. Ten *BnRLPs* presented expression specificity between R and S when elicited by *Sclerotinia sclerotiorum*, which might be fabulous candidates for *S. sclerotiorum* resistance research.

KEYWORDS

Brassica napus*, receptor-like protein, evolution, expression pattern, *Sclerotinia sclerotiorum

Introduction

Integrating exterior information with intrinsic cues is indispensable for all organisms, which is particularly crucial for plants due to their sessile lifestyle. Plants employ a large number and a vast variety of receptors for transducing signals from the extracellular matrix to the cell interior. Receptor-like proteins (RLPs), known as a kind of pattern recognition receptors (PRRs), are the main components of the first layer of plant immunity and a vital group of transmembrane receptors, which are generally composed of a short cytoplasmic domain, an extracellular leucine-rich repeat (LRR) domain, and a transmembrane domain (Wang et al., 2008). To date, 57 *At*RLPs (Wang et al., 2008), 90 RLPs in rice (Fritz-Laylin et al., 2005), 82 RLPs in poplar (Petre et al., 2014), 144 RLPs in cotton (Chen et al., 2015), 176 RLPs in tomato (Kang and Yeom, 2018), and 228 RLPs in *Brassica juncea* (Yang et al., 2020) have been identified, and pan-genome prediction of resistance gene analogs

(RGAs) in *Brassica oleracea* and also across the *Brassicaceae* were conducted (Bayer et al., 2019; Tirnaz et al., 2020). Currently, the enormous variation in RLPs between or within species has been discussed. RLPs are much more diverse and function in various pathways, involving both plant developmental regulation and immune response, even possessing dual roles. Indeed, most RLPs with assigned functions are involved in disease resistance. The identification of tomato Cf and Ve proteins that provide resistance against *Cladosporium fulvum* and *Verticillium spp*., respectively, started an era of RLP research (Jones et al., 1994; Kawchuk et al., 2001).

Subsequently, RLPs, for example, apple HcrVfs that confer resistance to fungal pathogens *Cladosporium fulvum* as well as *Venturia inaequalis* (Belfanti et al., 2004; Malnoy et al., 2008); tomato LeEIX proteins that function as a receptor to mediate recognition of the ethylene-inducing xylanase of *Trichoderma viride* (Ron and Avni, 2004); *Nicotiana benthamiana* elicitor-inducible leucine-rich repeat receptor-like protein (EILP) that is an ortholog of tomato Cf protein involved in bamboo mosaic virus movement (Chen et al., 2017); and Gbvdr6 that was homologous to Ve conferring resistance to *Verticillium dahliae Kleb* through the regulation of the JA/ET and SA signaling pathways in upland cotton were identified successively (Yang et al., 2017).

In *Arabidopsis*, of the 57 putative RLPs, approximately a quarter of them have been functionally characterized. CLAVATA2 (CLV2)/*At*RLP10 and TOO MANY MOUTHS (TMM)/*At*RLP17 are the first two *At*RLPs that are experimentally validated implicating in the development functions and are involved in regulating the meristem and organ development (Jeong et al., 1999), as well as stomatal patterning and distribution (Nadeau and Sack, 2002), respectively. FASCIATED EAR2, an ortholog of CLV2 in maize, also plays an important role in the regulation of stem cell homeostasis (Taguchi-Shiobara et al., 2001). It is noteworthy that functional redundancy exists among *At*RLPs. *At*RLP2 and *At*RLP12, identified as homologs of CLV2, rescue the *clv2* meristem defects when driven by the endogenous promoter of *CLV2* (Wang et al., 2010). Besides, the overexpression of *AtRLP3* and *AtRLP11* also rescued the phenotype of the *clv2-1* mutant (Wu et al., 2016a). Nevertheless, Steidele and Stam clustered *At*RLPs into two superclades (basal RLPs and pathogen-responsive RLPs) and referred to nine *At*RLPs (*At*RLP4, 10/CLV2, 17/TMM, 29, 44, 46, 51, 55, and 57) as putative developmental orthologs (PDOs) based on differences in transcript and protein abundance or clustering at the genomic loci (Steidele and Stam, 2021). Interestingly, CLV2/*At*RLP10 plays a role both in developmental and defense-related processes.

Over the years, more *At*RLPs have been shown to preferentially fulfill a role in pathogen defense. Among them, six *At*RLPs (RLP1, 3, 23, 30, 32, and 42) were validated as pathogen-responsive RLPs (Steidele and Stam, 2021). *At*RLP1/ReMAX is required for the perception of eMAX, which is a kind of MAMPs from *Xanthomonas* (Jehle et al., 2013a,b). *At*RLP3/RFO implicates in resistance to the vascular wilt fungus *Fusarium oxysporum* (Shen and Diener, 2013). *At*RLP23 perceives a conserved fragment found in most necrosis and ethylene-inducing peptide1-like proteins (NLPs) and thereby activates the immune responses, which are involved in pre-invasive resistance to the pathogen *Botrytis cinerea* (Albert et al., 2015; Ono et al., 2020). *At*RLP30 and *At*RLP18 impact the susceptibility of *Pseudomonas syringae* pv. *phaseolicola* to non-host resistance (Wang et al., 2008). Additionally, *At*RLP30 involves resistance toward necrotrophic fungal pathogen *Sclerotinia sclerotiorum* as well (Zhang et al., 2013). *At*RLP32 is confirmed as the receptor of proteobacterial translation initiation factor 1 (IF1) (Fan et al., 2022). *At*RLP42/RBPG1 recognizes endopolygalacturonases from the plant pathogen *Botrytis cinerea* and saprotroph *Aspergillus niger* (Zhang et al., 2014). *At*RLP44 activates brassinosteroid signaling through interaction with BAK1 (Wolf et al., 2014). *At*RLP51/SNC2 and *At*RLP55 implicate in basal defense against the bacterial pathogen *Pseudomonas syringae* pv *tomato* DC3000 (Zhang et al., 2010). *At*RLP52 is suggested to influence resistance toward the powdery mildew pathogen *Erysiphe cichoracearum* (Ramonell et al., 2005). Previous studies indicated that disease resistance genes are more likely to be duplicated and underwent functional divergence compared to growth-related genes, and the most homologous *AtRLP* genes anchored at the same or adjacent locus and underwent massive duplications (Fritz-Laylin et al., 2005; Wang et al., 2008; Wu et al., 2016a). Generally, RLPs need to constitutively interact with additional components, such as RLKs, to activate cellular responses. For example, SUPPRESSOR OF BAK1-INTERACTING RECEPTOR-LIKE KINASE 1 (SOBIR1) is required for RLP-mediated resistance to bacterial, fungal, and oomycete pathogens. SOBIR1 physically interacts with Cf-4, Ve1, *At*RLP1, *At*RLP23, *At*RLP30, *At*RLP32, and *At*RLP42, indicating the existence of crosstalk among different RLP signaling pathways (Wang et al., 2008; Jehle et al., 2013a; Liebrand et al., 2013; Zhang et al., 2013, 2014; Albert et al., 2015; Ma and Borhan, 2015; Catanzariti et al., 2017). Apart from the involvement in biotic stress responses, RLP was also engaged in abiotic stress tolerance. Currently, *At*RLP28 is the only one found to play a role in salt stress tolerance (Wu et al., 2016a).

Brassica napus is one of the most economically important oil crops in the world. Its yield and quality, however, are affected remarkably by multiple external stimuli, especially the invasion of major *Brassica* pathogenic fungi, such as *S. sclerotiorum* and *Leptosphaeria maculans*. *Brassica napus* (AACC) are an allotetraploid generated by recombination of two diploid genomes, *Brassica rapa* (AA) and *B. oleracea* (CC). *Brassica napus* and its progenitor species share significant homology with *Arabidopsis thaliana* (Nagaharu, 1935). Currently, several pathogen-responsive RLPs were identified across *Brassica* species. Among them, *LepR3* and *Rlm2*, which co-localized within the same genetic interval of blackleg resistance locus,

were cloned from *B. napus* and rendered race-specific resistance to fungal pathogen *L. maculans* upon recognition of different cognate Avr proteins, AvrLm1 and AvrLm2, respectively, and both of them require SOBIR1 for their functions (Larkan et al., 2013, 2015; Ma and Borhan, 2015). Another two *RLP* genes identified from *B. rapa* are *Bra032747* and *Bra032746*, which confer resistance to *Brassica* downy mildew (Zhang et al., 2018). Pathogen-associated cell-surface receptors are important for perceiving immunogenic signals in the challenged host (Kourelis and van der Hoorn, 2018). However, only a limited batch of *RLPs* was functionally characterized in *Brassica* species. Here, we performed the identification and phylogenetic analysis of *RLP*s in *B. napus* and its progenitors, *B. rapa* and *B. oleracea*, which are highly homologous to *At*RLPs, and sorted out their distributions on chromosomes. To further assume the possible functions of identified *RLPs*, we investigated gene expression profiles of *B. napus* elicited by *S. sclerotiorum* utilizing the published pathogen-induced RNA-seq datasets. Taken together, these results will sketch out the basic information across identified *Cruciferae RLPs*, contributing to deciphering genome evolution and duplication, as well as elucidation of *RLP* gene function in detail.

Materials and methods

Identification of *RLP* genes

Arabidopsis thaliana genomic and annotation data were downloaded from the TAIR (http://www.arabidopsis.org). Genomic and annotation data of *B. napus* and *B. rapa* were downloaded from BRAD (http://brassicadb.cn/; Chen et al., 2022). *Brassica oleracea* genomic and annotation data were downloaded from Ensemblplants (https://plants.ensembl.org/index.html).

In *A. thaliana*, a typical RLP protein consists of tandem LRR motifs, along with SP (signal peptide) and TM (transmembrane) in two terminals (Wang et al., 2008). But these LRR motifs vary in number and distribution, thus they are not competent as a good query for protein screening. Given this, two methods were used for the identification of RLP-encoding genes in this work. First, as queries, 55 identified *Arabidopsis* RLP protein sequences (two pseudogenes, *AtRLP18* and *AtRLP49*, excluded in this work) were obtained and used to search analogs in three *Brassica* species using the BLASTP program based on a comparative genomics approach, thereby obtaining the first part of the candidate RLP proteins. Furthermore, RLP-encoding genes were also screened through Hidden Markov Model (HMM) profiles built by the model corresponding to 55 *At*RLP proteins using the HMMSearch program (https://www.ebi.ac.uk/Tools/hmmer/). Subsequently, the two sets of candidates were merged and subjected to NCBI-BLASTP using UniProtKB/SwissoPrpt, and their annotations were downloaded for further screening.

Gene structure, conserved motif, and phylogenetic analyses

The exon–intron organization of the RLP-encoding genes was depicted using GSDS2.0 (Hu et al., 2015), and the conserved motifs in RLPs were searched using the MEME suite (Bailey and Elkan, 1994). The SP and TM were predicted using SignalP-5.0 Server (http://www.cbs.dtu.dk/services/SignalP/), SMART (http://smart.embl-heidelberg.de/), TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), and Pfam (http://pfam.xfam.org/). Approximately 2,000-bp upstream flanking fragments of the RLP genes were derived from the genome and used for promoter cis-element prediction using PlantCARE (Lescot et al., 2002). The visualization was completed using TBtools (Chen et al., 2020). The phylogenetic analyses of the selected *B. napus* and *Arabidopsis* RLP proteins were generated using MEGA7.0 with the maximum likelihood (ML) algorithm. Bootstrap analysis with 1,000 replications was performed to assess group support.

Physical location of *RLP*s on the chromosome

The chromosome location of *Brassicaceae RLP*s was obtained from the genome annotation files. The distribution and tandem duplication of *Brassicaceae RLP* genes on chromosomes was generated and depicted using TBtools (Chen et al., 2020).

Expression characteristics of *BnRLP* genes induced by pathogens

The RNA-seq data were obtained from NCBI. Data calibrations were performed, and some data with big deviations were removed. A heat map was drawn by HemI 1.0.3.3 (Deng et al., 2014). A mid-resistant *B.napus* variety Zhongyou821 was used for *S. sclerotiorum* inoculation and assessment of *BnRLP* gene expression. Plants were grown in a greenhouse with artificial irrigation. At the flowering stage, stems and leaves were inoculated with mycelia on living plants, harvested after 48 h inoculation, flash-frozen in liquid nitrogen, and stored at −80°C. Three plants with the closest phenotype and growth status were harvested, and harvesting was repeated three independent times. qRT-PCR was performed as described above with slight modifications (Li et al., 2016). Total RNA samples were isolated from rapeseed tissues using the Plant RNAprep Pure Kit (Tiangen, Beijing). RNA was quantified on a NanoDrop 1000 (NanoDrop Technologies, Inc.), and RNA integrity was evaluated on a 1% agarose gel. RNA was transcribed into cDNA using a GoScript RT Reagent Kit (Promega, USA). Primers used for qRT-PCR were designed using the Primer Premier

5.0 program to target the ORF of each gene with an amplicon sized between 80 and 200 bp (Supplementary Table 3). UBC9 served as a reference gene (Chen et al., 2010). qRT-PCR was performed using 10-fold diluted cDNA and a Universal SYBR Green Supermix Kit (Bio-RAD, USA) on a CFX96 real-time PCR machine (Bio-Rad, USA) according to the MIQE Guidelines (Minimum Information for Publication of Quantitative Real-Time PCR Experiments; Bustin et al., 2009).

Results

Identification and characterization of RLP-encoding genes in *B. napus*

Since *A. thaliana* is a close relative of *B. napus* (rapeseed), 55 *At*RLP protein sequences (two pseudogenes, *At*RLP18 and *At*RLP49, excluded in this work) were downloaded from the *Arabidopsis* database (TAIR) and used as queries to identify the candidate RLP proteins in rapeseed by BLAST search against BRAD or Ensemblplants database. In addition, Hidden Markov Model (HMM) was used for further confirmation of candidate RLPs. To better understand the evolution of RLPs, *B. rapa* and *B. oleracea* genomes were also selected for the screening of orthologs. In total, 118, 173, and 276 RLPs were identified from *B. rapa* (AA), *B. oleracea* (CC), and *B. napus* (AACC), respectively (Supplementary Table 1). Although a pangenome investigation of resistance gene analogs (RGAs) across some *B. oleracea* varieties had identified 213 candidates (Bayer et al., 2019), we conducted a further fine selection to get a more reliable but attenuated set of BoRLPs. Then, all *RLP* CDSs were downloaded and aligned to corresponding genomic DNA to obtain high-confidence sequences. Sequence alignment revealed a great diversity of *BnRLP*s (Supplementary Figure 1). The number of introns ranged from 0 to 22 (*BnaC06g23240D*), and ∼16% *BnRLP*s had no introns which were sharply attenuated compared to that of *AtRLP*s (70%). The orthologs of five *AtRLP*s (*AtRLP3*, *AtRLP10*, *AtRLP17*, *AtRLP44,* and *AtRLP51*) did not evolve any intron, showing high conservation. Apart from quantitative variation of *BnRLP* introns, some *BnRLP* genes possessed extra-long introns (more than 2,000 bp), like *BnaA01g31040D* (intron I, 5,336 bp), *BnaA09g35430D* (intron II, 4,756 bp; intron III, 5,018,bp), *BnaC02g22740D* (intron III, 6,474bp), and *BnaC02g22760D* (intron X, 5,030 bp). *AtRLP9* (*AT*1G58190.1) also contained long introns (3,437 bp), but it had no corresponding orthologs in *Brassicaceae* species. Rapeseed amplified the RLP-encoding gene copies by a wide margin, and the tendency of increasing intron number and length variation was available for alternative splicing and further benefit for functional diversity.

Further analysis revealed that the composition and the number of conserved motifs or domains varied to a large extent. The length of amino acids ranged from 73 (*Bna*Cnng19280D) to 2,736 aa (*Bna*C06g23240D; Supplementary Figure 2). Ten motifs were predicted in *Bn*RLP proteins (Figure 1, Supplementary Figure 3). Motif 1, Motif 2, and Motif 9 were distributed on more than 80% *Bn*RLP proteins, and the average number on each protein was 2.3, 1.7, and 2.2, respectively. More than 15% of *Bn*RLP proteins possessed all 10 motifs, and 42 of the 276 had less than five types. Only a quarter of *Bn*RLP proteins met the characteristics of canonical RLP protein, namely composed of the SP and TM domains at the ends and the tandem LRR domains in the middle (Supplementary Figure 1).

Chromosomal localization and gene expansion

BnRLP encoding genes were unevenly distributed on eighteen chromosomes (Figure 2). The 116 and 148 *BnRLP*s were mapped onto chrA (Figure 2A) and chrC (Figure 2B) subgenome, respectively, and many of these genes reside in a cluster manner. The percentage of *RLP* genes on chromosomes in clusters in *B. oleracea* (51.22%) and *B. rapa* (49.14%) was lower than that of *B. napus* (56.06%). The numbers of genes in clusters ranged from two to six in *B. napus*, and the maximum gene number in *B. oleracea* and *B. rapa* were 7 and 5, respectively. In *B. napus*, 148 RLP genes were located in 55 clusters and the remaining 221 genes were singletons. Among these clusters, 12 with 29 genes were located on chrC04 (Figure 2B), which was similar to that in *B. oleracea* (seven clusters containing 23 genes on chromosome C04) (Figure 2D). The *B. rapa* genome carries 57 RLP genes in 23 clusters (Figure 2C). Compared to that in *B.napus*, the same number of clusters but fewer RLP genes (57 and 67 in *B. rapa* genome and *B.napus* A-subgenome) were observed in *B. rapa*. The cluster distribution between the two species was obviously different on chromosome A01 and chrA01, showing that *B.napus* evolved more tandem *RLP* genes on chrA01. Chromosome chrC04 contained the largest number of *Bn*RLP genes (41 genes), followed by chrC03 (21 genes), chrC06 (19 genes), chrA04 (18 genes), and chrA01 (14 genes). Only one *BnRLP* gene was located on chrA10. *RLP* genes were assigned tandemly on most chromosomes, especially on chrA01, chrA04, chrA05, chrA07, chrC04, and chrC06, which presented higher distribution densities. Generally, the distribution of *BnRLP* genes on the A or C subgenome was similar to that of *BrRLP* or *BoRLP* genes in their respective genomes.

To better reveal the expansion of *RLP* genes in the rapeseed genome, the duplication patterns of 276 *BnRLP* genes were predicted and analyzed by MCScanX (Wang et al., 2012) and visualized by TBtools (Chen et al., 2020), and the result showed that rapeseed had 36 times as many *RLP* gene pairs as *Arabidopsis* (Figure 3). Moreover, the possible syntenic

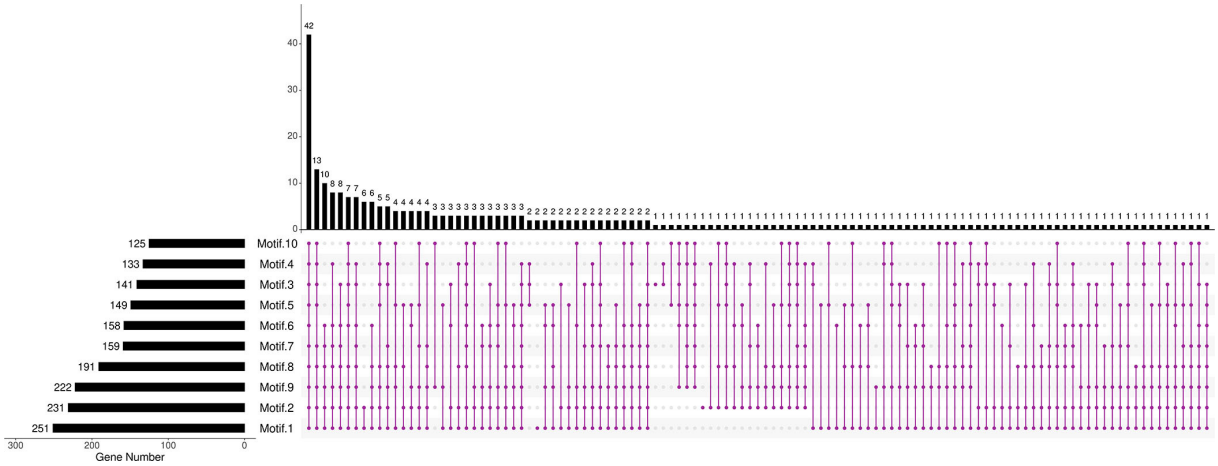


FIGURE 1
Distributions of motifs in *Bn*RLPs. The Upset plot shows the distribution of motifs in the *Bn*RLPs. The number chart above represents the number of genes contained in each type of RLP. The bar chart at the bottom left represents the number of events included in each type of motif. The dotted line shows the types of motifs contained in the group.

relationship of RLP-encoding genes between *A. thaliana* and *Brassica* genomes was also investigated. Subsequently, we obtained 59 orthologous gene pairs between *A. thaliana* and *B. oleracea*, 50 between *A. thaliana* and *B. rapa*, 250 between *B. oleracea* and *B. napus*, and 223 between *B. rapa* and *B. napus*, which are shown in Supplementary Figure 4. Ka/Ks analysis was performed by TBtools (Chen et al., 2020) between *A. thaliana* and the other three *Brassica* species(*B. napus*, *B. rapa* and *B. oleracea*), respectively. The Ka/Ks value of all pairs was <1 (Supplementary Table 2), suggesting that the main force for the evolution of those RLP gene pairs was negative selection. Most orthologous *AtRLP* genes are located on chromosomes 1 and 2, and correspondingly, chromosomes A04/A07 and C04/C06 possessed the higher orthologous regions in *B. rapa* and *B. oleracea*, respectively. Approximately, the aliquot of orthologous *BnRLP* genes was resourced from *B. rapa* and *B. oleracea* and numerically distributed evenly on chrA and chrC subgenomes.

Evolution analysis of *Bn*RLPs

To better understand the evolving relationship between *Bn*RLPs and *At*RLPs, a phylogenetic tree was constructed by the ML method (Figure 4). The RLP proteins were clustered into 13 subgroups. Group I to Group IV consisted of no more than three proteins, respectively. The protein constructs of most *Bn*RLPs in GroupV, X, and XI were canonical. More than half of *Bn*RLPs in GroupVI had an SP domain or a TM domain. Most *Bn*RLPs possessed only one type of domain, and two TM domains were predicted in four *Bn*RLPs in GroupXII. In addition, besides a sub-clade of *Bn*RLPs that only possessed the TM domain (one or more), the quantity of *Bn*RLPs that the missing TM domain had increased to 69.23% in GroupXII, but most had an SP domain. Some *RLP* genes in GroupVII were more conserved since most members had no introns. Some bunches of tandem *AtRLPs* were especially prone to cluster together, which evolved independently with *BnRLPs* in Group XI and Group XIII. Other tandem ones evolved together with *BnRLPs*, like *At1g17240* (*AtRLP2*) and *At1g17250* (*AtRLP3*) in GroupVII, *At2g32660* (*AtRLP22*) and *At2g32680* (*AtRLP23*) in GroupXIII.

Analysis of *cis*-acting elements in the promoter region of *BnRLPs*

Cis-acting elements that are distributed in the promoter region can help predict the function of candidate genes. Most *At*RLPs and *Bn*RLPs were identified as related to stress responses, so the *cis*-acting elements related to stresses were further analyzed (Supplementary Figure 5, Figure 5). Concretely, *cis*-acting elements were categorized into MeJA-responsive (MeJA, name after), gibberellin-responsive (GA), abscisic acid-responsive (ABA), drought-inducible (drought), auxin-responsive (auxin), low temperature-responsive (low temperature), salicylic acid-responsive (SA), wound-responsive (wound), and defense- and stress-responsive (defense and stress) related. Approximately, 79.3 and 77.5% of the promoters contained MeJA and ABA-related *cis*-acting elements, followed by GA- and wound-related, which were found in half of the promoters of *BnRLPs*. Only 15 *BnRLPs* were related to wound responsiveness. Additionally, a quarter contained no more than three types of *cis*-acting elements. Among them, seven *BnRLPs* might exhibit functional specificity since they possessed single type of *cis*-acting element, such as *BnaA09g37810D*

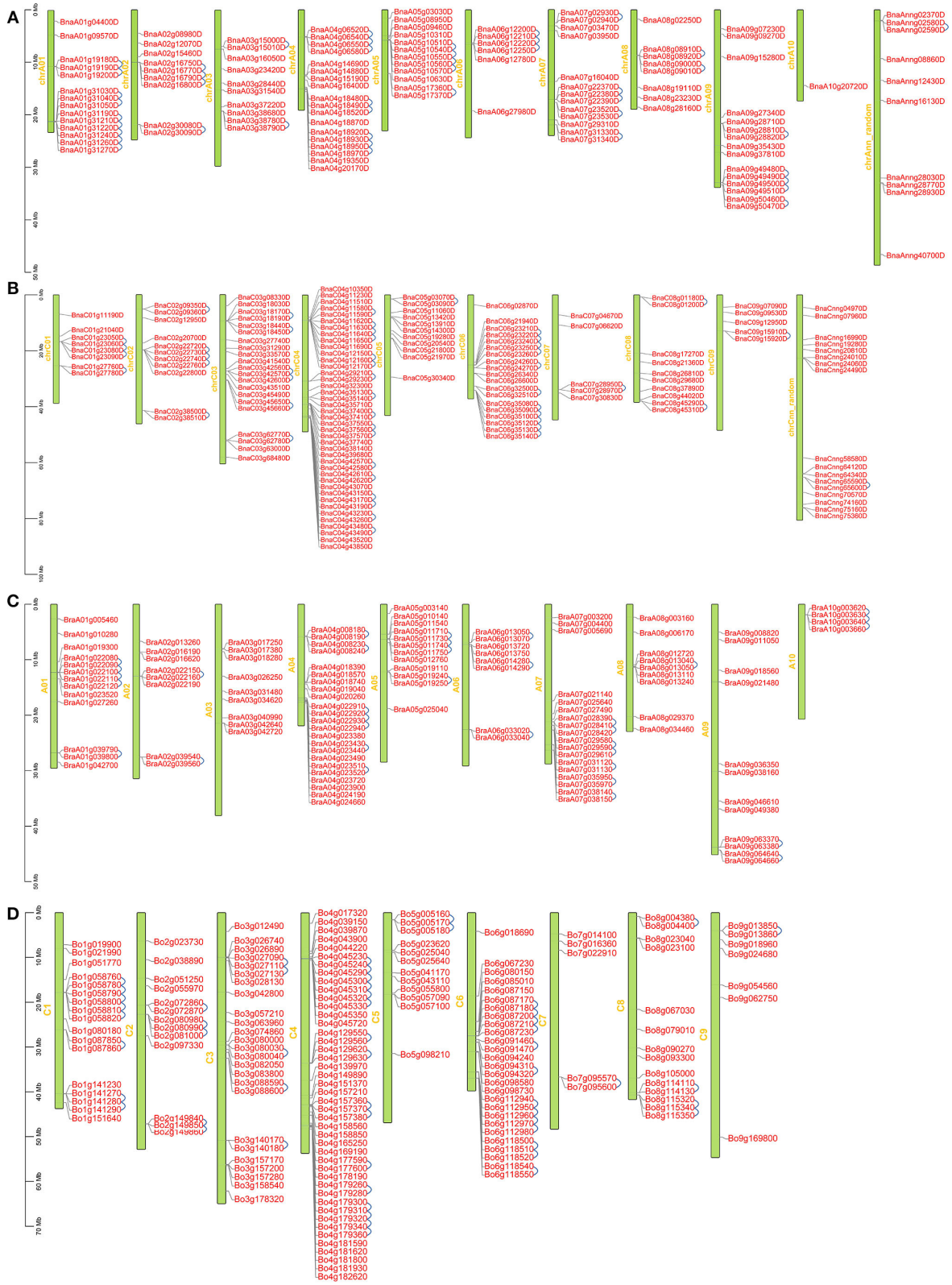


FIGURE 2
Locations of *RLP* genes on chromosomes in *Brassica napus*, *Brassica rapa* and *Brassica oleracea*. **(A, B)** Locations of *RLP* genes on chromosomes in *B. napus*; **(C)** Locations of *RLP* genes on chromosomes in *B. rapa*. **(D)** Locations of *RLP* genes on chromosomes in *B. oleracea*. A01-A09, C01-C08, and chrA01-chrC09 represent the chromosome number in *B. rapa*, *B. oleracea*, and *B. napus*, respectively. The blue line represents tandem duplication.

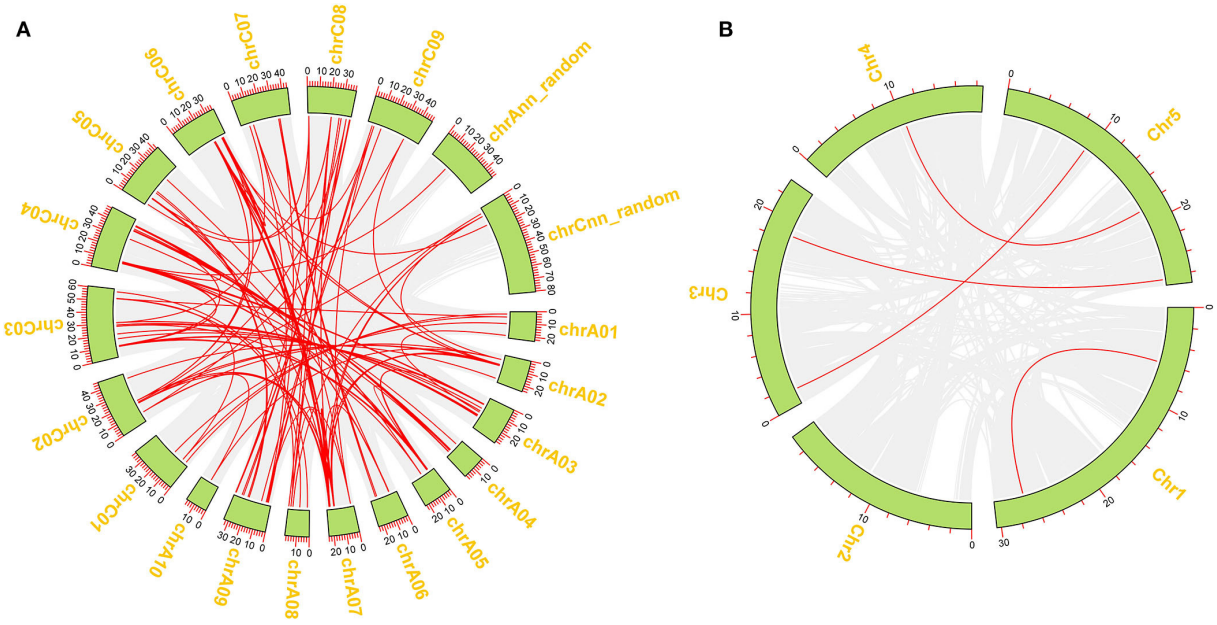


FIGURE 3
Duplication of *BnRLP* and *AtRLP* genes. **(A)** Duplication of *RLP* genes in *Brassica napus*. **(B)** Duplication of *RLP* genes in *Arabidopsis thaliana*. The red line represents the gene pairs and the green box represents the genome.

(drought), *BnaA02g12070D* (SA), *BnaC02g20700D* and *BnaC09g12950D* (MeJA), *BnaC02g45850D* and *BnaC04g12150D* (low temperature), *BnaC09g15910D* (defense and stress), and *BnaC04g52590D* (ABA).

Expression pattern of *RLP* genes induced *by Sclerotinia sclerotiorum*

Sclerotinia sclerotiorum is one of the most common pathogens worldwide, which is the causal agent of stem rot disease in *Brassica* crops causing devastating decline in the economic value. To further unveil the protagonists in pathogen-defense responses, expression profiling data of *RLP* genes were derived from the GEO database. One series (accession No. SRP053361) (Wu et al., 2016b) was downloaded and analyzed. In total, 225 *BnRLP* genes identified showed expressional fluctuations during 24, 48, and 96 hpi (hour post-inoculated) by *S. sclerotiorum* in resistance (R) or susceptible (S) rapeseeds (Figure 6). To confirm the reliability of the dataset and further identify the genes that can be induced by *S. sclerotiorum*, seven upregulated *BnRLP* gene expressions were detected using real-time fluorescent quantitative PCR (Supplementary Figure 6). As we observed, besides inflorescence, the plant stem and leaves were also susceptible to *S. sclerotiorum,* and the infected stem becomes more prone to lodging at the flowering stage, so the fact that the RLPs are induced at a higher expression level at the flowering stage might have the potential to possess greater agricultural applicability. To this end, the stem and leaves were selected as target tissues to illuminate the tissue specificity under *S. sclerotiorum* infection. The results showed that those *RLP* genes can be induced significantly by *S. sclerotiorum*, but varied in expression level and pattern. Among them, *BnaC07g28970D*, *BnaC02g22760D*, and *BnaA02g16770D* accumulated more transcripts in the stem than that in the leaf, and the differences ranged from 2- to 4-fold. On the contrary, *BnaA08g02250D*, *BnaA08g28160D*, *BnaC04g56380D*, and *BnaA04g14880D* were induced more in the leaf. Moreover, *BnaA03g56320D* showed a similar expression level in stem and leaf, but at a relatively lower level compared to other genes. Overall, the results showed that all selected upregulated genes derived from the database could be induced by *S. sclerotiorum.*

Unexpectedly, only 29 *BnRLPs* were upregulated (more than two times compared to mock) at a certain time point (Figure 7). Among them, some *BnRLPs* presented expression specificity between R and S, such as four *BnRLPs* (*BnaA04g14880D*, *BnaA05g08950D*, *BnaA04g14690D*, *BnaC09g15920D*) that were induced merely in S, and six *BnRLPs* (*BnaA09g37810D*, *BnaA03g56320D*, *BnaA05g10570D*, *BnaC06g02870D*, *BnaC03g45660D*, and *BnaC04g37400D*) in R. The expressional differences of some *BnRLPs* between R and S appeared due to the inconsistency of the induction time. For instance, transcript accumulations of *BnaC07g28970D*, *BnaC02g22760D*, and *BnaA02g16770D* in R were slower than S, and their expression

FIGURE 4
Phylogenetic trees of RLP proteins in *Arabidopsis thaliana* and *Brassica napus*. The phylogenetic tree was generated by the ML method with bootstrap analysis (1,000 bootstrap replicates) using an amino acid sequence alignment of RLP proteins from *A. thaliana* and *B. napus* by the MEGA 7.0 program. RLP proteins were clustered into 13 subgroups. The rings surrounded by the phylogenetic tree stand for two types of domains. The inner one represents the transmembrane domain and the outer represents the signal peptide. "Y" means that RLP contains one transmembrane domain/signal peptide, "YY" for two transmembrane domains, and "YYYYYY" for six transmembrane domains. "N" means that RLP does not contain transmembrane domain/signal peptide.

levels at 96 hpi (R) were approximately equal to that at 48 hpi (S). Additionally, as results show, *BnaC02g22760D* and *BnaA02g16770D* were induced from 24hpi in S, and their expression levels continued increasing even at 96 hpi reaching a level of 13.85 and 31.03 times, respectively, but they were induced from 96 hpi in R. *BnaA08g28160D* and *BnaC04g56380D* possessed similar expression patterns between R and S. The transcript accumulations of *BnaA08g28160D* reached a peak at 48 hpi in both R and S. Furthermore, *BnaC04g56380D* was induced at 48 hpi and kept accumulating until 96 hpi in two rapeseed ecotypes. It is worth mentioning that *LepR3*/*Rlm2* encodes a receptor-like protein triggered by the *L. maculans* effector AVRLM1 (Larkan et al., 2013, 2015), but the expression of *BnaA10g20720D*, which possessed 100% identity and was co-located in the same genetic interval with *LepR3*/*Rlm2*, was inhibited by *S. sclerotiorum* both in R and S (Figure 6).

Discussion

The fluctuation of gene expression profiles induced by various pathogens can be used as an indicator for functional

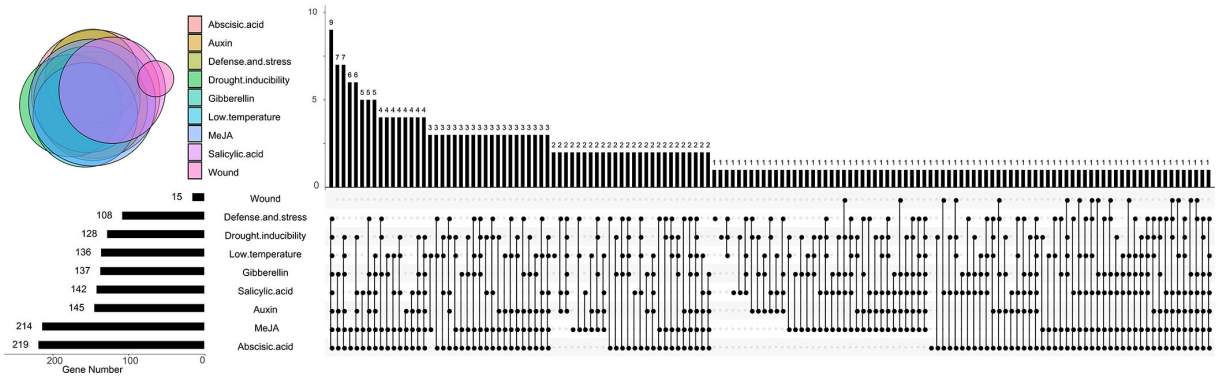


FIGURE 5
Distributions of *cis*-acting elements in *BnRLP* promoter. The Upset plot shows the distribution of *cis*-acting elements in *BnRLP* promoter. The bar chart above represents the number of genes contained in each type of RLP. The bar chart at the bottom left represents the number of genes included in each type of *cis*-acting element. The dotted line shows the types of *cis*-acting elements contained in the group.

assumption and further characterization. To investigate the possible functions of *BnRLP*s, we compared the transcriptional profiles of *RLP* genes with that of the functional assigned *AtRLP* genes. Intriguingly, *AtRLP10* (*CLV2, AT1G65380.1*) was identified as a regulator of plant meristem and organ development (Jeong et al., 1999; Wang et al., 2011). As its homologs, the transcript abundance of *BnaC02g45200D* and *BnaA02g12070D* were decreased when there was a stimulus exerted by *S. sclerotiorum*. The expressional inclination of *RLP10* genes indicated that their functions might be conserved among *Brassica* species and *A. thaliana* due to the trade-off between growth and defense response, which is widely known in plants. Another hint in favor of the possibility of functional conservation among *RLP10* genes is that they shared high similarity/identity with *AtRLP10* no matter in sequence (CDS and protein) or structural layer (Supplementary Figure 7). Unlike *RLP10*, a functional divergence occurred among the *RLP30* genes as *AtRLP30* (*AT3G05360.1*) was suggested to be involved in resistance toward *S. sclerotiorum*, while *BnRLP30* genes (*BnaAnng02580D* and *BnaAnng02370D*) were not induced by it (Figure 6). The low similarity/identity between *BnRLP30s* and *AtRLP30* also supported the probability of functional divergence in *B. napus* (Supplementary Figure 7). Instead, *BnaA02g16770D* and *BnaC02g22760D*, homologs of *AtRLP15* (*AT1G74190.1*), might be fabulous candidates for *S. sclerotiorum* resistance research since they were proven to be few but induced to a greater extent.

The case that one single *RLP* gene is transcriptionally regulated by multiple external stimuli and involved in multiple biological processes has been reported in *A. thaliana* (Wu et al., 2016a). It was also a common phenomenon in *B. napus,* as exemplified by *BnaA09g09270D*, which can be induced by both *S. sclerotiorum* and *L. maculans* (Becker et al., 2017; Duke et al., 2017). The comparative analysis of *BnRLP*s phylogeny and their transcriptional profiles indicated that there was no significant correlation between *BnRLP*s phylogeny (predicted by sequence similarity) and the expression patterns, and this is in line with previous publications (Steidele and Stam, 2021). On the contrary, the diversification of gene expression patterns was observed among the duplicated *BnRLP*s paralogs. Gene duplication widely exists during the process of plant evolution in multiple forms, like tandem duplication, segmental duplication, and whole-genome duplication (Kondrashov et al., 2002; Conant and Wolfe, 2008). Retained duplicated genes can provide raw genetic resources for functional innovation, and the novel function can derive from several different ways, including gain-of-function mutations, the subdivision of ancestral functions, and gene dosage changes (Conant and Wolfe, 2008). In this study, tandem duplication was commonly found in *BnRLP*s, and some *BnRLP*s were amplified by a wide margin when compared with their counterparts in *A. thaliana*, such as homologs of *AtRLP27* (*AT2G33060.1*) and *AtRLP54* (*AT5G40170.1*).

As *B. napus* was hybridized by *B. rapa* and *B. oleracea*, in general, the evolution of BnRLP proteins was relatively conserved among the three *Brassica* species. However, because of gene duplications and translocations after hybridization, inaccurate assembly universally occurred during genetic recombination in plants, resulting in the loss-of-function of some genes (Schnable et al., 2012; Boutte et al., 2020). In this study, 5, 10, and 24 proteins were detected as containing no LRR domains in *B. rapa*, *B. oleracea*, and *B. napus*, respectively, but remained as RLPs. They possessed a high sequence identity and had a similar protein construct (consisted of SP and/or TM domain) with other typical RLPs. So those proteins with no LRR domains were also included to better reveal the evolution of BnRLPs.

When we did the phylogenetic analysis, the phylogeny of *At*RLPs resembles that performed by previous publications

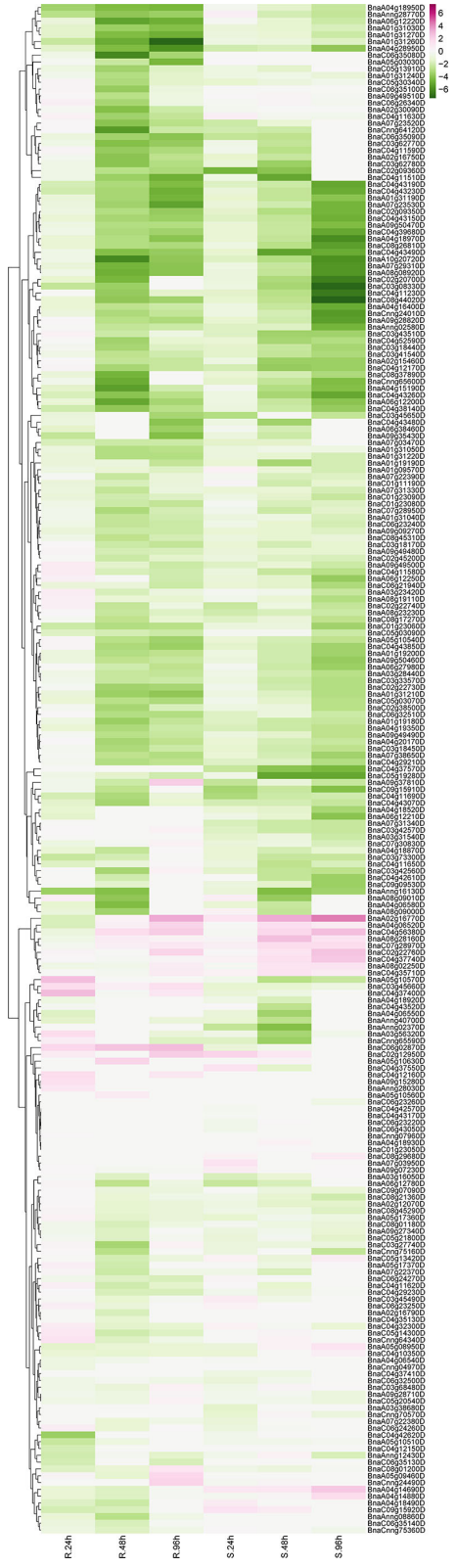


FIGURE 6
The expression heat map of *BnRLP* genes elicited by *S. sclerotiorum*. The expression heat map of *BnRLP* genes in the *(Continued)*

FIGURE 6
resistant (R) and susceptible (S) *Brassica napus* alleles at 24, 48, and 96 h post-inoculation, TPM value derived from the GEO database was used for heat map construction.

(Wang et al., 2008; Steidele and Stam, 2021). In line with the classification by Steidele and Stam (2021), most upregulated *Bn*RLPs were homologous with *At*RLPs that were clustered into pathogen-responsive RLP clade. The exception is that *BnaA02g16770D* and *BnaC02g22760D*, homologs of *AtRLP15* (*AT1G74190.1*) which were grouped into a basal sub-family but not the aforementioned pathogen-responsive RLP clade, upregulated after pathogen treatment to a great extent.

Based on the dataset, six *Bn*RLP genes, namely *BnaA09g37810D*, *BnaA03g56320D*, *BnaA05g10570D*, *BnaC06g02870D*, *BnaC03g45660D*, and *BnaC04g37400D*, upregulated specifically in resistant rapeseed by *S. sclerotiorum* elicitation. Correspondingly, four *Bn*RLPs (*BnaA04g14880D*, *BnaA05g08950D*, *BnaA04g14690D*, and *BnaC09g15920D*) showed preference for susceptible rapeseed. Those genes identified from incompatible interactions might be good candidates for further functional assignment of RLP genes against *S. sclerotiorum* infection. *BnaA02g16770D* and *BnaC02g22760D,* which could be induced universally by *S. sclerotiorum*, might also play a role in preventing fungal infection. Zhongyou821 was identified as a mid-resistant *B.napus* variety (Wang et al., 2004) and commonly used as a control cultivar for investigating disease indexes. In this study, we also used Zhongyou821 for infection and performed qRT-PCR to verify the expression level of some upregulated *BnRLP* genes, and the result showed that the transcript abundance in this study was dozens of times than that derived from the RNAseq datasets, but further confirmed they could be induced by *S. sclerotiorum.* The large difference might be due to the selection of different cultivars or ascribed to spatiotemporal expression differences. Overall, all selected upregulated *BnRLPs* genes derived from the database could be induced by *S. sclerotiorum* in Zhongyou821.

Conclusion

We conducted identification, gene and protein characterization, localization, evolution, and expression analysis of the highly conserved RLP members in *B. napus* and found that a majority of RLP gene families in this study were relatively conserved during the evolution of *Brassicaceae.* Exploring effective resistance genes against a vast variety of microbial pathogens is critical for crop breeding. Genome-wide identification and expression analysis of the RLP family members in *B. napus* provided an alternative strategy to

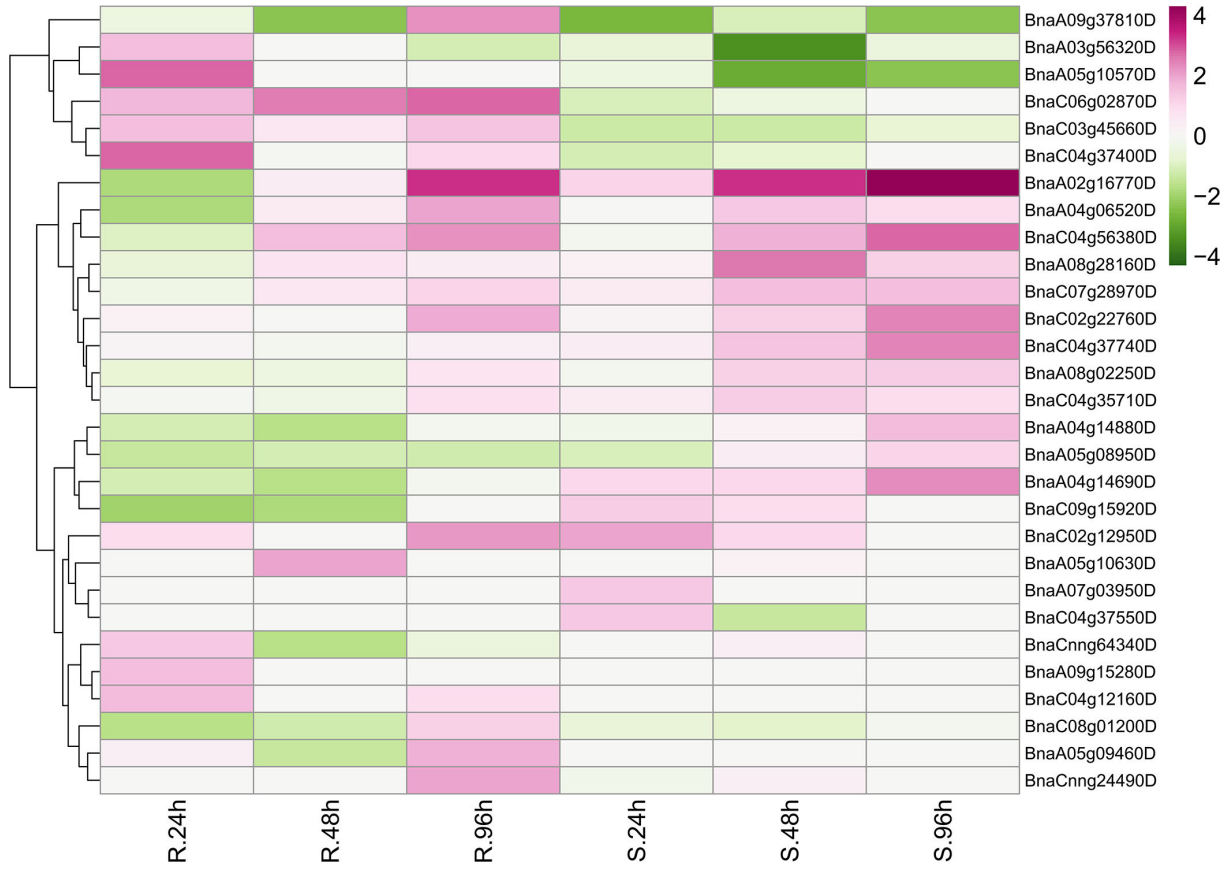


FIGURE 7
Upregulated *BnRLP*s elicited by *Sclerotinia sclerotiorum*. The expression heat map of upregulated *BnRLP* genes in the resistant (R) and susceptible (S) *B. napus* alleles at 24, 48, and 96 h post-inoculation by *S. sclerotiorum*, TPM value derived from the GEO database was used for heat map construction.

reinforce the resistance against major pathogens in *Brassicaceae*. Our results provide important clues for further investigations of the function of *Brassicaceae* RLPs involved in the development and immune response and pave the way to molecular breeding of disease-resistant rapeseed.

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

WL: formal analysis, writing the original draft, data curation, and validation. JL: investigation, resources, and visualization. CY: data curation. SX: conceptualization, writing—review and editing, and supervision. 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 (Grants 31971836 and 30970247), the Hunan Province Natural Science Fund for Distinguished Young Scholars (Grant 11JJ1007), the Chongqing municipal education commission (Grant KJQN201900533), and the Chongqing Bureau of Human Resources and Social Security Post-doctoral Funding (0019).

Acknowledgments

We thank Yanan Liu for the critical reading of the manuscript. We thank all members of our laboratories for their helpful assistance during the research.

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.2022.944763/full#supplementary-material

SUPPLEMENTARY TABLE 1
RLP gene information in *B. napus*, *B. rapa* and *B. oleracea*.

SUPPLEMENTARY TABLE 2
Ka/Ks in *A.thaliana* and *B. napus*, *B. rapa* and *B. oleracea* pairs.

SUPPLEMENTARY TABLE 3
Primers in this study.

SUPPLEMENTARY FIGURE 1
Gene and protein structures of *RLP* genes in *B. napus*. The green bar represents CDS, the black line represents intron, and the blue bar represents UTR. The pink bar stands for the LRR motif, the yellow bar for the transmembrane domain, and the red for the signal peptide.

SUPPLEMENTARY FIGURE 2
Motif and conserved domain of RLP in *B. napus*. Motif labeled as motif 1-motif 10. The green bar represents CDS, the blue bar represents UTR. The yellow bar stands for the conserved domain PLN03150 and the pink bar for PLN00113.

SUPPLEMENTARY FIGURE 3
Motifs in *Bn*RLPs.

SUPPLEMENTARY FIGURE 4
Syntenic relationship of RLP-encoding genes between *A. thaliana* and *Brassicaceae* species.

SUPPLEMENTARY FIGURE 5
Cis-acting element in *BnRLP* promoter.

SUPPLEMENTARY FIGURE 6
Tissue specificity of upregulated *BnRLP*s under *S. sclerotiorum* infection. The box plot represents dispersion and expression degree. Gene expression was detected with at least three biological repeats and three experiment repeats, and statistical analysis showed that compared with controls, the expression of all genes reached significant levels when elicited by *S. sclerotiorum* (p-value < 0.05).

SUPPLEMENTARY FIGURE 7
Protein similarity and identity of RLP families. Data on the upper right present the protein sequence identity, and data on the bottom left present the protein sequence similarity.

References

Albert, I., Böhm, H., Albert, M., Feiler, C. E., Imkampe, J., Wallmeroth, N., et al. (2015). An RLP23-SOBIR1-BAK1 complex mediates NLP-triggered immunity. *Nat Plants* 1, 15140. doi: 10.1038/nplants.2015.140

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Bayer, P. E., Golicz, A. A., Tirnaz, S., Chan, C.-K. K., Edwards, D., and Batley, J. (2019). Variation in abundance of predicted resistance genes in the *Brassica oleracea* pangenome. *Plant Biotechnol. J.* 17, 789–800. doi: 10.1111/pbi.13015

Becker, M. G., Zhang, X., Walker, P. L., Wan, J. C., Millar, J. L., Khan, D., et al. (2017). Transcriptome analysis of the *Brassica napus-Leptosphaeria maculans* pathosystem identifies receptor, signaling and structural genes underlying plant resistance. *Plant J.* 90, 573–586. doi: 10.1111/tpj.13514

Belfanti, E., Silfverberg-Dilworth, E., Tartarini, S., Patocchi, A., Barbieri, M., Zhu, J., et al. (2004). The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proc. Natl. Acad. Sci. USA.* 101, 886–890. doi: 10.1073/pnas.0304808101

Boutte, J., Maillet, L., Chaussepied, T., Letort, S., Aury, J.-M., Belser, C., et al. (2020). Genome size variation and comparative genomics reveal intraspecific diversity in *Brassica rapa*. *Front. Plant Sci.* 11, 577536. doi: 10.3389/fpls.2020.577536

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797

Catanzariti, A.-M., Do, H. T. T., Bru, P., de Sain, M., Thatcher, L. F., Rep, M., et al. (2017). The tomato I gene for *Fusarium wilt* resistance encodes an atypical leucine-rich repeat receptor-like protein whose function is nevertheless dependent on SOBIR1 and SERK3/BAK1. *Plant J.* 89, 1195–1209. doi: 10.1111/tpj.13458

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

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

Chen, I.-H., Huang, Y.-P., Tseng, C.-H., Ni, J.-T., Tsai, C.-H., Hsu, Y.-H., et al. (2017). *Nicotiana benthamiana* elicitor-inducible leucine-rich repeat receptor-like protein assists bamboo mosaic virus cell-to-cell movement. *Front. Plant Sci.* 8, 1736. doi: 10.3389/fpls.2017.01736

Chen, J.-Y., Huang, J.-Q., Li, N.-Y., Ma, X.-F., Wang, J.-L., Liu, C., et al. (2015). Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium wilt*. *BMC Plant Biol.* 15, 148. doi: 10.1186/s12870-015-0508-3

Chen, X., Truksa, M., Shah, S., and Weselake, R. J. (2010). A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in *Brassica napus*. *Anal. Biochem.* 405, 138–140. doi: 10.1016/j.ab.2010.05.032

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Deng, W., Wang, Y., Liu, Z., Cheng, H., and Xue, Y. (2014). HemI: a toolkit for illustrating heatmaps. *PLoS ONE* 9, e111988. doi: 10.1371/journal.pone.0111988

Duke, K. A., Becker, M. G., Girard, I. J., Millar, J. L., Dilantha Fernando, W. G., Belmonte, M. F., et al. (2017). The biocontrol agent *Pseudomonas chlororaphis*

PA23 primes *Brassica napus* defenses through distinct gene networks. *BMC Genomics* 18, 467. doi: 10.1186/s12864-017-3848-6

Fan, L., Fröhlich, K., Melzer, E., Pruitt, R. N., Albert, I., Zhang, L., et al. (2022). Genotyping-by-sequencing-based identification of *Arabidopsis* pattern recognition receptor RLP32 recognizing proteobacterial translation initiation factor IF1. *Nat. Commun.* 13, 1294. doi: 10.1038/s41467-022-28887-4

Fritz-Laylin, L. K., Krishnamurthy, N., Tör, M., Sjölander, K. V., and Jones, J. D. G. (2005). Phylogenomic analysis of the receptor-like proteins of rice and *Arabidopsis*. *Plant Physiol.* 138, 611–623. doi: 10.1104/pp.104.054452

Hu, B., Jin, J., Guo, A., 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

Jehle, A. K., Fürst, U., Lipschis, M., Albert, M., and Felix, G. (2013a). Perception of the novel MAMP eMax from different *Xanthomonas* species requires the *Arabidopsis* receptor-like protein ReMAX and the receptor kinase SOBIR. *Plant Signal. Behav.* 8, e27408. doi: 10.4161/psb.27408

Jehle, A. K., Lipschis, M., Albert, M., Fallahzadeh-Mamaghani, V., Fürst, U., Mueller, K., et al. (2013b). The receptor-like protein ReMAX of *Arabidopsis* detects the microbe-associated molecular pattern eMax from *Xanthomonas*. *Plant Cell* 25, 2330–2340. doi: 10.1105/tpc.113.110833

Jeong, S., Trotochaud, A. E., and Clark, S. E. (1999). The *Arabidopsis* CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell* 11, 1925–1934. doi: 10.1105/tpc.11.10.1925

Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., and Jones, J. D. (1994). Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* 266, 789–793. doi: 10.1126/science.7973631

Kang, W.-H., and Yeom, S.-I. (2018). Genome-wide identification, classification, and expression analysis of the receptor-like protein family in tomato. *Plant Pathol. J.* 34, 435–444. doi: 10.5423/PPJ.OA.02.2018.0032

Kawchuk, L. M., Hachey, J., Lynch, D. R., Kulcsar, F., van Rooijen, G., Waterer, D. R., et al. (2001). Tomato Ve disease resistance genes encode cell surface-like receptors. *Proc. Natl. Acad. Sci. USA.* 98, 6511–6515. doi: 10.1073/pnas.091114198

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biol.* 3, RESEARCH0008. doi: 10.1186/gb-2002-3-2-research0008

Kourelis, J., and van der Hoorn, R. A. L. (2018). Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579

Larkan, N. J., Lydiate, D. J., Parkin, I., A,. P., Nelson, M. N., Epp, D. J., et al. (2013). The *Brassica napus* blackleg resistance gene LepR3 encodes a receptor-like protein triggered by the *Leptosphaeria maculans* effector AVRLM1. *New Phytol.* 197, 595–605. doi: 10.1111/nph.12043

Larkan, N. J., Ma, L., and Borhan, M. H. (2015). The *Brassica napus* receptor-like protein RLM2 is encoded by a second allele of the LepR3/Rlm2 blackleg resistance locus. *Plant Biotechnol. J.* 13, 983–992. doi: 10.1111/pbi.12341

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de 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, 325–327. doi: 10.1093/nar/30.1.325

Li, W., Lu, J., Lu, K., Yuan, J., Huang, J., Du, H., et al. (2016). Cloning and Phylogenetic analysis of *Brassica napus* L. caffeic acid o-methyltransferase 1 gene family and its expression pattern under drought stress. *PLoS ONE* 11, e0165975. doi: 10.1371/journal.pone.0165975

Liebrand, T. W. H., van den Berg, G. C. M., Zhang, Z., Smit, P., Cordewener, J. H. G., America, A. H. P., et al. (2013). Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. *Proc. Natl. Acad. Sci. USA.* 110, 10010–10015. doi: 10.1073/pnas.1220015110

Ma, L., and Borhan, M. H. (2015). The receptor-like kinase SOBIR1 interacts with *Brassica napus* LepR3 and is required for *Leptosphaeria maculans* AvrLm1-triggered immunity. *Front. Plant Sci.* 6, 933. doi: 10.3389/fpls.2015.00933

Malnoy, M., Xu, M., Borejsza-Wysocka, E., Korban, S. S., and Aldwinckle, H. S. (2008). Two receptor-like genes, Vfa1 and Vfa2, confer resistance to the fungal pathogen *Venturia inaequalis* inciting apple scab disease. *Mol. Plant Microbe Interact.* 21, 448–458. doi: 10.1094/MPMI-21-4-0448

Nadeau, J. A., and Sack, F. D. (2002). Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* 296, 1697–1700. doi: 10.1126/science.1069596

Nagaharu, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* 7, 389–452.

Ono, E., Mise, K., and Takano, Y. (2020). RLP23 is required for *Arabidopsis* immunity against the grey mould pathogen *Botrytis cinerea*. *Sci. Rep.* 10, 13798. doi: 10.1038/s41598-020-70485-1

Petre, B., Hacquard, S., Duplessis, S., and Rouhier, N. (2014). Genome analysis of poplar LRR-RLP gene clusters reveals RISP, a defense-related gene coding a candidate endogenous peptide elicitor. *Front. Plant Sci.* 5, 111. doi: 10.3389/fpls.2014.00111

Ramonell, K., Berrocal-Lobo, M., Koh, S., Wan, J., Edwards, H., Stacey, G., et al. (2005). Loss-of-function mutations in chitin responsive genes show increased susceptibility to the powdery mildew pathogen *Erysiphe cichoracearum*. *Plant Physiol.* 138, 1027–1036. doi: 10.1104/pp.105.060947

Ron, M., and Avni, A. (2004). The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *Plant Cell* 16, 1604–1615. doi: 10.1105/tpc.022475

Schnable, J. C., Freeling, M., and Lyons, E. (2012). Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol. Evol.* 4, 265–277. doi: 10.1093/gbe/evs009

Shen, Y., and Diener, A. C. (2013). *Arabidopsis thaliana* resistance to fusarium oxysporum 2 implicates tyrosine-sulfated peptide signaling in susceptibility and resistance to root infection. *PLoS Genet.* 9, e1003525. doi: 10.1371/journal.pgen.1003525

Steidele, C. E., and Stam, R. (2021). Multi-omics approach highlights differences between RLP classes in *Arabidopsis thaliana*. *BMC Genomics* 22, 557. doi: 10.1186/s12864-021-07855-0

Taguchi-Shiobara, F., Yuan, Z., Hake, S., and Jackson, D. (2001). The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. *Genes Dev.* 15, 2755–2766. doi: 10.1101/gad.208501

Tirnaz, S., Bayer, P. E., Inturrisi, F., Zhang, F., Yang, H., Dolatabadian, A., et al. (2020). Resistance gene analogs in the brassicaceae: identification, characterization, distribution, and evolution. *Plant Physiol.* 184, 909–922. doi: 10.1104/pp.20.00835

Wang, G., Ellendorff, U., Kemp, B., Mansfield, J. W., Forsyth, A., Mitchell, K., et al. (2008). A genome-wide functional investigation into the roles of receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 147, 503–517. doi: 10.1104/pp.108.119487

Wang, G., Long, Y., Thomma, B. P. H. J., de Wit, P. J. G. M., Angenent, G. C., and Fiers, M. (2010). Functional analyses of the CLAVATA2-like proteins and their domains that contribute to CLAVATA2 specificity. *Plant Physiol.* 152, 320–331. doi: 10.1104/pp.109.148197

Wang, G., Zhang, Z., Angenent, G. C., and Fiers, M. (2011). New aspects of CLAVATA2, a versatile gene in the regulation of *Arabidopsis* development. *J. Plant Physiol.* 168, 403–407. doi: 10.1016/j.jplph.2010.08.015

Wang, H. Z., Liu, G. H., Zheng, Y. B., Wang, X. F., and Yang, Q. (2004). Breeding of the *Brassica napus* cultivar Zhongshuang 9 with high-resistance to *Sclerotinia sclerotiorum* and dynamics of its important defense enzyme activity. *Sci. Agric. Sin.* 37, 23–28.

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wolf, S., van der Does, D., Ladwig, F., Sticht, C., Kolbeck, A., Schürholz, A.-K., et al. (2014). A receptor-like protein mediates the response to pectin modification by activating brassinosteroid signaling. *Proc. Natl. Acad. Sci. USA.* 111, 15261–15266. doi: 10.1073/pnas.1322979111

Wu, J., Liu, Z., Zhang, Z., Lv, Y., Yang, N., Zhang, G., et al. (2016a). Transcriptional regulation of receptor-like protein genes by environmental stresses and hormones and their overexpression activities in *Arabidopsis thaliana*. *J. Exp. Bot.* 67, 3339–3351. doi: 10.1093/jxb/erw152

Wu, J., Zhao, Q., Yang, Q., Liu, H., Li, Q., Yi, X., et al. (2016b). Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Sci. Rep.* 6, 19007. doi: 10.1038/srep19007

Yang, H., Bayer, P. E., Tirnaz, S., Edwards, D., and Batley, J. (2020). Genome-wide identification and evolution of receptor-like kinases (RLKs) and receptor like proteins (RLPs) in *Brassica juncea*. *Biology* 10, 17. doi: 10.3390/biology10010017

Yang, Y., Chen, T., Ling, X., and Ma, Z. (2017). Gbvdr6, a gene encoding a receptor-like protein of cotton (*Gossypium barbadense*), confers resistance to verticillium wilt in *Arabidopsis* and upland cotton. *Front. Plant Sci.* 8, 2272. doi: 10.3389/fpls.2017.02272

Zhang, B., Li, P., Su, T., Li, P., Xin, X., Wang, W., et al. (2018). BrRLP48, encoding a receptor-like protein, involved in downy mildew resistance in *Brassica rapa*. *Front. Plant Sci.* 9, 1708. doi: 10.3389/fpls.2018.01708

Zhang, L., Kars, I., Essenstam, B., Liebrand, T. W. H., Wagemakers, L., Elberse, J., et al. (2014). Fungal endopolygalacturonases are recognized as microbe-associated molecular patterns by the *Arabidopsis* receptor-like protein RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiol.* 164, 352–364. doi: 10.1104/pp.113.230698

Zhang, W., Fraiture, M., Kolb, D., Löffelhardt, B., Desaki, Y., Boutrot, F. F. G., et al. (2013). *Arabidopsis* receptor-like protein30 and receptor-like kinase suppressor of BIR1-1/EVERSHED mediate innate immunity to necrotrophic fungi. *Plant Cell* 25, 4227–4241. doi: 10.1105/tpc.113.117010

Zhang, Y., Yang, Y., Fang, B., Gannon, P., Ding, P., Li, X., et al. (2010). *Arabidopsis* snc2-1D activates receptor-like protein-mediated immunity transduced through WRKY70. *Plant Cell* 22, 3153–3163. doi: 10.1105/tpc.110.074120

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fpls.2022.841328

Map-Based Cloning and Characterization of *Br-dyp1*, a Gene Conferring Dark Yellow Petal Color Trait in Chinese Cabbage (*Brassica rapa* L. ssp. *pekinensis*)

Shuangjuan Yang[1,2†], Honglei Liu[1,2†], Yanyan Zhao[1†], Henan Su[1], Xiaochun Wei[1], Zhiyong Wang[1], Xiaobin Zhao[1], Xiao-Wei Zhang[1,2] and Yuxiang Yuan[1,2*]*

[1]Institute of Horticulture, Henan Academy of Agricultural Sciences, Zhengzhou, China, [2]School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China

Flower color is an important trait in *Brassica* species. However, genes responsible for the dark yellow flower trait in Chinese cabbage have not been reported. In this study, we identified a dark-yellow-flowered Chinese cabbage line SD369. Genetic analysis indicated that the dark yellow flower trait in SD369 was controlled by a single recessive locus, *Br-dyp1* (*dark yellow petal color* 1 in *Brassica rapa*). Using bulked segregant RNA sequencing and kompetitive allele-specific PCR assays, *Br-dyp1* was fine-mapped to an interval of 53.6 kb on chromosome A09. Functional annotation analysis, expression analysis, and sequence variation analysis revealed that *Bra037130* (*BraA09.ZEP*), which encodes a zeaxanthin epoxidase, was the most likely candidate gene for *Br-dyp1*. Carotenoid profile analysis suggested that *Bra037130* (*BraA09.ZEP*) might participate in the epoxidation from zeaxanthin to violaxanthin. The 679 bp insertion in dark yellow petal caused premature stop codon, thus caused the loss-of-function of the enzyme zeaxanthin epoxidase (ZEP), which disturbed the carotenoid metabolism, and caused the increased accumulation of total carotenoid, and finally converted the flower color from yellow to dark yellow. Comparative transcriptome analysis also showed that the "carotenoid biosynthesis" pathway was significantly enriched, and genes involved in carotenoid degradation and abscisic acid biosynthesis and metabolism were significantly downregulated. Furthermore, we developed and validated the functional marker Br-dyp1-InDel for *Br-dyp1*. Overall, these results provide insight into the molecular basis of carotenoid-based flower coloration in *B. rapa* and reveal valuable information for marker-assisted selection breeding in Chinese cabbage.

Keywords: ***Brassica rapa*, flower color, dark yellow petal color, fine-mapping, breeding**

OPEN ACCESS

Edited by:
Keiichi Okazaki,
Niigata University, Japan

Reviewed by:
Pengfang Zhu,
Shenyang Agricultural University, China
Zhansheng Li,
Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Xiao-Wei Zhang
xiaowei5737@163.com
Yuxiang Yuan
yuxiangyuan126@126.com

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

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

Received: *22 December 2021*
Accepted: *27 January 2022*
Published: *17 February 2022*

Citation:
Yang S, Liu H, Zhao Y, Su H, Wei X, Wang Z, Zhao X, Zhang X-W and Yuan Y (2022) Map-Based Cloning and Characterization of Br-dyp1, a Gene Conferring Dark Yellow Petal Color Trait in Chinese Cabbage (Brassica rapa L. ssp. pekinensis). Front. Plant Sci. 13:841328. doi: 10.3389/fpls.2022.841328

INTRODUCTION

Flower color is one of the most important traits in *Brassica* species and is particularly useful for ornamental and landscaping purposes (Zhang et al., 2018a; Liu et al., 2020b). In breeding, flower color can be used to evaluate variety purity in hybrid production (Zhang et al., 2018b; Yang et al., 2021). *Brassica* flowers are usually yellow, but can also be white, milky white, orange, or dark yellow (Zhang et al., 2018a). The gene *carotenoid cleavage dioxygenase 4* (*CCD4*) underlies the white flower trait in *Brassica napus* and *Brassica oleracea* (Zhang et al., 2015; Han et al., 2019). In *Brassica rapa*, the orange flower color as well as the orange coloration of the inner leaves is controlled by the *carotenoid isomerase* (*BrCRTISO*) gene (Su et al., 2014). Our previous study showed that the *BrWF3* gene, which encodes a diacylglycerol acyltransferase and is homologous to *PES2* in *Arabidopsis*, controls the white flower trait in Chinese cabbage (*B. rapa* L. ssp. *pekinensis*; Yang et al., 2021). However, the gene underlying the dark yellow flower trait in Chinese cabbage or in *B. rapa* species has not been reported.

Carotenoids are a group of more than 700 lipid-soluble pigments synthesized in plastids (Gonzalez-Jorge et al., 2016; Li et al., 2016). In chloroplasts, carotenoids are essential structural and functional components of the antenna complex of the photosynthesis system, with roles in light harvesting, non-photochemical quenching, and limiting membrane damage by reactive oxygen species and singlet oxygen species (Walter and Strack, 2011; Zhang et al., 2015; Gonzalez-Jorge et al., 2016). Carotenoids in chromoplasts endow flowers and fruits with different colors, such as orange, yellow, and red, which attract animals for pollination and seed dispersal (Kevan and Baker, 1983). Carotenoids are also biosynthetic precursors for the synthesis of the plant hormones abscisic acid (ABA) and strigolactone (Frey et al., 2006; Alder et al., 2012).

The enzyme zeaxanthin epoxidase (ZEP) plays a critical role in carotenoid biosynthesis, which is responsible for the epoxidation of zeaxanthin to yield antheraxanthin and subsequently violaxanthin. Violaxanthin can be deepoxidated to antheraxanthin and then zeaxanthin by the enzyme violaxanthin de-epoxidase. This reversible epoxidation/deepoxidation is referred to as the xanthophyll cycle, in which deepoxidation to zeaxanthin is favored under high-light conditions while epoxidation to violaxanthin predominates in moderate-light conditions. The rapid formation of zeaxanthin *via* the xanthophyll cycle is indispensable for the dissipation of excess energy by non-photochemical quenching, while violaxanthin is a precursor for ABA biosynthesis; thus, xanthophyll cycle is one of the critical processes contributing to plant fitness and stress tolerance (Alboresi et al., 2011; Gao et al., 2013; Gonzalez-Jorge et al., 2016; Lou et al., 2017). Many studies have showed that ZEP proteins are involved in ABA biosynthesis, *ZEP* mutants display low ABA levels and almost no ABA upregulation under drought stress, while the overexpression of *ZEP* enhances tolerance to osmotic stress, which suggests that the ZEP enzyme plays a critical role in the ABA-mediated stress response (Marin et al., 1996; Agrawal et al., 2001; Xiong et al., 2002; Park et al., 2008). However, the impact of ZEP on carotenoid pigmentation in the flowers of Chinese cabbage has not been investigated.

In this study, we conducted positional cloning of the dark yellow petal color gene (*Br-dyp1*) in Chinese cabbage by using F_2 populations, which derived from the dark-yellow-flowered inbred line "SD369" and the yellow-flowered DH line "R16-11." Furthermore, we conducted carotenoid profile analysis and comparative transcriptome analysis to figure out the mechanisms underlying the dark yellow flower color pigmentation. In addition, we developed and validated a functional marker. This work will promote marker-assisted selection breeding and the exploration of molecular mechanisms that regulate flower color variation in Chinese cabbage or in *B. rapa*.

MATERIALS AND METHODS

Plant Materials

The yellow-petaled double haploid (DH) line R16-11 (P_1) and the dark-yellow-petaled inbred Chinese cabbage line SD369 (P_2) were used as parents to generate F_1, F_2, BC_1P_1, and BC_1P_2 populations for the inheritance and mapping studies. The BC_1P_1 and BC_1P_2 populations were created by backcrossing F_1 plants with R16-11 or SD369, respectively. The petal color trait was investigated visually at the flowering stage. Statistical analyses of the segregation ratios of the F_2 and BC_1P_1 populations were carried out through chi-square test (χ^2). Additionally, eight yellow-petaled materials, eight dark-yellow-petaled materials, eight orange-petaled materials, and eight white-petaled materials were used to analyze mutations in the candidate gene (**Supplementary Table 1**). All the materials used in this study were provided by the Institute of Horticulture, Henan Academy of Agricultural Sciences.

Carotenoids Identification and Quantification

Carotenoids composition was measured by MetWare[1] based on the AB Sciex QTRAP 6500 LC–MS/MS platform. Petals from 10 dark-yellow-petaled F_2 plants were combined to form one replicate (referred to as the DY-bulk), and petals from 10 yellow-petaled F_2 plants were included in the Y-bulk. In total, three replicates were assessed. Fresh petals were freeze-dried, ground into powder (30 Hz, 1.5 min), and stored at −80°C until needed. For each sample, 50 mg powder was weighted and extracted with 0.5 ml of a mixed solution of n-hexane: acetone: ethanol (1:1:1, v/v/v) with 0.01% BHT (g/ml), and 10 μl of (13C10)-β-carotene solution (20 μg/ml) were added into the extract as internal standards for quantification. The extract was vortexed for 20 min at room temperature. The supernatants were collected after centrifugation at 12000 r/min for 5 min at 4°C. The residue was re-extracted by repeating the above steps again under the same conditions. Saturated sodium chloride solution (0.5 ml) was added to the supernatant, after which the mixture was vortexed, and the upper layer

[1]http://www.metware.cn/

was collected. This step was repeated two times more. Then, the supernatant was evaporated to dryness and dissolved in 0.5 ml of MTBE, then 0.5 ml 10% KOH-MeOH was added, the mixture was vortexed again, and the reaction was allowed to take place at room temperature overnight. After the reaction, 1 ml of saturated sodium chloride solution and 0.5 ml of MTBE were added, followed by vortexing, and the upper layer was collected. This step was repeated two times, and the supernatant was evaporated to dryness and reconstituted in 100 µl of mixed solution of MeOH/MTBE (1:1, v/v). The solution was filtered through a 0.22 µm membrane filter for further LC–MS/MS analysis.

The sample extracts were analyzed using a UPLC-APCI-MS/MS system (UPLC, ExionLC™ AD; MS, Applied Biosystems 6,500 Triple Quadrupole), which was performed as described in previous studies (Liu et al., 2020a; Wang et al., 2020; Zhou et al., 2020; Yang et al., 2021). The integrated peak area of each carotenoid detected in the samples was substituted into the linear equations of standard curves for content calculation; finally, the absolute content data for the carotenoids in the actual samples were obtained (**Supplementary Table 2**). The specific procedure for calculation of the carotenoid content was performed as Yang et al. (2021).

Bulked Segregant RNA Sequencing and Analysis

The DY-bulk and Y-bulk each with three replicates used for carotenoid analysis were also used for RNA sequencing. Six cDNA were constructed and sequenced at BioMarker Tech Co., Ltd. (Beijing, China). The Illumina HiSeq X 10 platform was used to generate 150-base paired-end reads for each library. To preliminarily map the candidate gene, the clean reads from the three DY-bulk replicates were merged to form a single read file, and another merged file was obtained from the three Y-bulk replicates. Then, the merged read files were aligned to the *B. rapa* reference genome (V1.5) using BWA software (Li and Durbin, 2010). The single-nucleotide polymorphism (SNP) and insertion/deletion (InDel) variants were called using SAMtools software (Li et al., 2009). The SNP index was calculated for all genomic positions in the DY-bulk and Y-bulk and the Δ(SNP index) was calculated by subtracting the SNP index of the Y-bulk from that of the DY-bulk, which was performed as previously described (Abe et al., 2012; Takagi et al., 2013; Yang et al., 2021). The candidate region of *Br-dyp1* was identified by sliding window analysis with a 1-Mb width and a 50-kb increment at the 95% confidence level according to (Yang et al., 2021).

To identify differentially expressed genes (DEGs) between the DY-bulk and Y-bulk, the clean reads of each library were aligned to the *B. rapa* V1.5 reference genome using HISAT2 software with the default parameters (Kim et al., 2015). Then, the fragments per kilobase of transcript per million mapped reads value of each gene was calculated to estimate gene expression levels. DEGs were identified using the DESeq2 package (v1.6.3; Love et al., 2014). Genes with false discovery rate (FDR) ≤ 0.05 and $|\log2$ (fold change) $| \geq 1$ were recognized as DEGs. Gene Ontology (GO) enrichment analysis was carried out using the topGO package (v2.18.0; Alexa et al., 2010). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was implemented using KOBAS (v2.0) software (Mao et al., 2005; Wu et al., 2006).

Kompetitive Allele-Specific PCR Marker and Linkage Map Development

To validate the results of Bulked Segregant RNA Sequencing (BSR-Seq) and map the *Br-dyp1* gene, we selected SNPs showing polymorphism between the two bulks and nearing the candidate region for Kompetitive Allele-Specific PCR (KASP) marker development. The detailed procedures for KASP marker development and KASP assays were performed as described by (Yang et al., 2020). The developed KASP markers were first screened between R16-11 and SD369. Then, polymorphic KASP markers (**Supplementary Table 3**) were employed to genotype the F_2 population containing 94 individuals. The genetic linkage map was constructed using JoinMap 4.0 software (Van Ooijen, 2006). Recombination values were converted into genetic map distances (cM) following the Kosambi mapping function (Kosambi, 1943).

For the fine-mapping of the candidate gene, approximately 3,500 F_2 seeds were planted, and 743 individuals with a dark yellow petal phenotype were used for screening more recombinants.

Cloning and Sequence Analysis of the Candidate Genes

We designed primer pairs according to the *B. rapa* reference genome to clone the DNA and cDNA sequence of the candidate gene. Phanta Flash Master Mix (Vazyme Biotech Co., Ltd., Nanjing, China) was deployed to amplify the candidate gene. The PCR products were sequenced by Sunya Biotech Co., Ltd. (Zhengzhou, China), and sequence alignments were performed using DNAMAN software. The coding sequences (CDSs) of candidate gene from R16-11 and SD369 were submitted to GenBank under the accession numbers OL436220 (R16-11) and OL436221 (SD369).

Quantitative Real-Time PCR

Total RNA was extracted from tissues of roots, stems, leaves, sepals, petals, stamens, and pistils from R16-11 and SD369. *BrGAPDH* was used as an internal control (Qi et al., 2010; Su et al., 2014). Quantitative Real-Time PCR (qRT-PCR) was performed with 2 × TB Green Premix Ex Taq II (TaKaRa, Japan) on a Roche LightCycler 480-II System (Roche Applied Sciences, Beijing, China). Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The results from three biological replicates are shown.

RESULTS

The Dark Yellow Petal Color Trait of SD369 Is Controlled by a Single Recessive Gene

The phenotypic analysis showed significant differences in petal color between the two parental lines (**Figure 1**). In R16-11 (P_1), the petals showed stable yellow coloration at the flowering stage, whereas those of SD369 (P_2) exhibited dark yellow

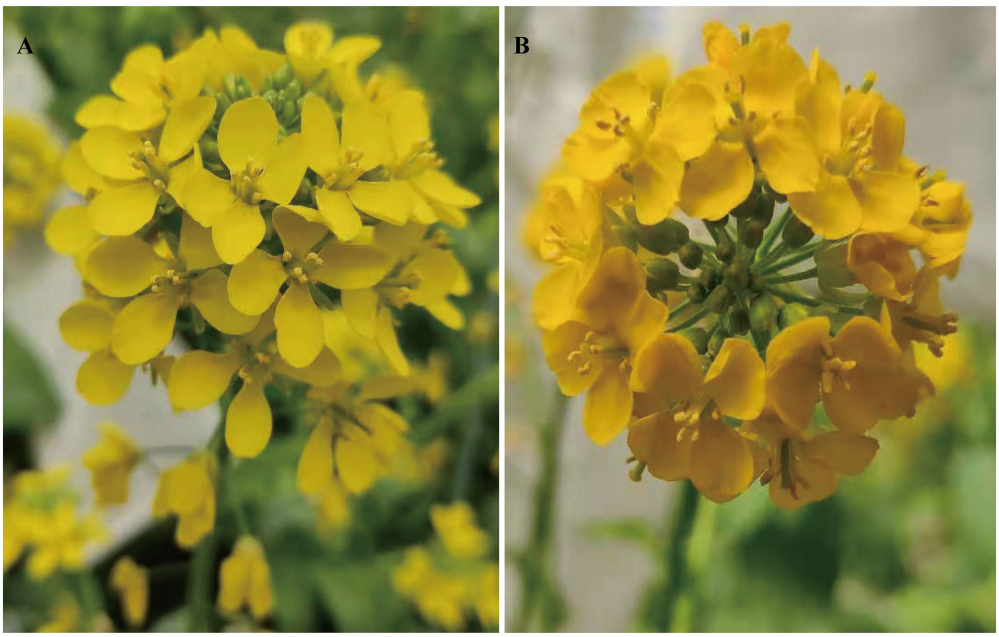


FIGURE 1 | Phenotypic characterization of flower color in the two parent lines (R16-11 and SD369). The flower color of R16-11 **(A)** is yellow at the flowering stage, while dark yellow for SD369 **(B)**.

coloration (**Figures 1A,B**). All 15 individuals in the F_1 population showed a yellow petal phenotype, as did R16-11. Phenotypic segregation was observed in the F_2 and BC_1P_2 populations, and the petals of the plants in these two populations exhibited two kinds of colorations, yellow or dark yellow, corresponding to the coloration of either R16-11 or SD369, respectively. In a small F_2 population, 176 plants exhibited yellow petals, and 54 showed dark yellow petals, corresponding to a segregation ratio of 3:1 by the chi-square test (**Table 1**). In a larger F_2 population, the segregation ratio was also 3:1 (2,371 yellow:743 dark yellow, $\chi^2 = 2.16$). The numbers of individuals with yellow and dark yellow petals in the BC_1P_2 population were 125 and 115, respectively, corresponding to a ratio of 1:1 according to the chi-square test ($\chi^2 = 0.42 < \chi^2_{0.05} = 3.84$, $p < 0.05$). Furthermore, all 200 BC_1P_1 plants showed yellow petals. All of these results demonstrate that the dark yellow petal color trait of SD369 is controlled by a monogenic recessive gene (**Table 1**). We named this locus *Br-dyp1* (*dark yellow petal color 1 in B. rapa*).

The Carotenoid Profile Is Altered in Dark Yellow Petals

The composition and content of carotenoids in petals of the Y-bulk and DY-bulk were determined using UPLC-APCI-MS/MS system under saponification treatment. The results showed that the carotenoid profiles of the DY-bulk were quite different from those of the Y-bulk. Ten carotenoids composition were identified in Y-bulk and DY-bulk, including three carotenes and seven xanthophylls (**Figure 2**; **Supplementary Table 4**). The total content of carotenes was slightly lower in the DY-bulk than in the Y-bulk (**Supplementary Table 4**). The total content of xanthophylls accounted for approximately 92.0% or 98.9% of the total carotenoids in the Y-bulk and DY-bulk, respectively, and the contents of almost all xanthophylls were higher in the DY-bulk than in the Y-bulk. For example, the amounts of lutein, zeaxanthin, and antheraxanthin in the DY-bulk increased about 6.7-fold, 11.2-fold, and 11.5-fold than that in Y-bulk. However, the amount of violaxanthin decreased about 3.6-fold in DY-bulk (**Figure 2**; **Supplementary Table 4**). The total content of carotenoids showed a remarkable 4.7-fold increase in dark yellow petals relative to that in the yellow petals (**Supplementary Table 4**). Taken together, these findings suggested that the dark yellow petal phenotype in SD369 resulted from increased contents of carotenoids, particularly lutein, zeaxanthin, and antheraxanthin.

Fine-Mapping of the *Br-dyp1* Gene

Br-dyp1 was preliminarily mapped using BSR-seq. A total of 47,206,290, 38,634,054, and 54,612,412 clean reads were obtained for the three Y-bulks, while 47,676,208, 51,940,890, and 49,967,262 clean reads were obtained for the three DY-bulks (**Supplementary Table 5**). Reads from the three Y-bulks were merged as the Y-pool, and reads from the three DY-bulks were merged to form the DY-pool. The Y-pool and DY-pool clean reads were aligned to the *B. rapa* V1.5 genome, and a total of 348,456 SNPs and 29,092 InDels were identified between these two pools. The Δ(SNP index) of each position was calculated for sliding window analysis. According to the null hypothesis, a 2.7 Mb region from 3.8 to 6.5 Mb on chromosome A09 exhibiting significant linkage disequilibrium was identified as the candidate region for the dark yellow petal trait at a 99% confidence level

TABLE 1 | Genetic analysis of the petal trait in parents and in crosses between R16-11 and SD369.

Generations	Total	Yellow	Dark yellow	Expected ratio	χ^2	$\chi^2_{0.05}$
P_1 (R16-11)	10	10	0	–	–	–
P_2 (SD369)	10	0	10	–	–	–
F_1	15	15	0	–	–	–
F_2-small	230	176	54	3:1	0.28	3.84
F_2-large	3,114	2,371	743	3:1	2.16	3.84
BC_1P_1 (F_1 × R16-11)	200	200	0	–	–	–
BC_1P_2 (F_1 × SD369)	240	125	115	1:1	0.42	3.84

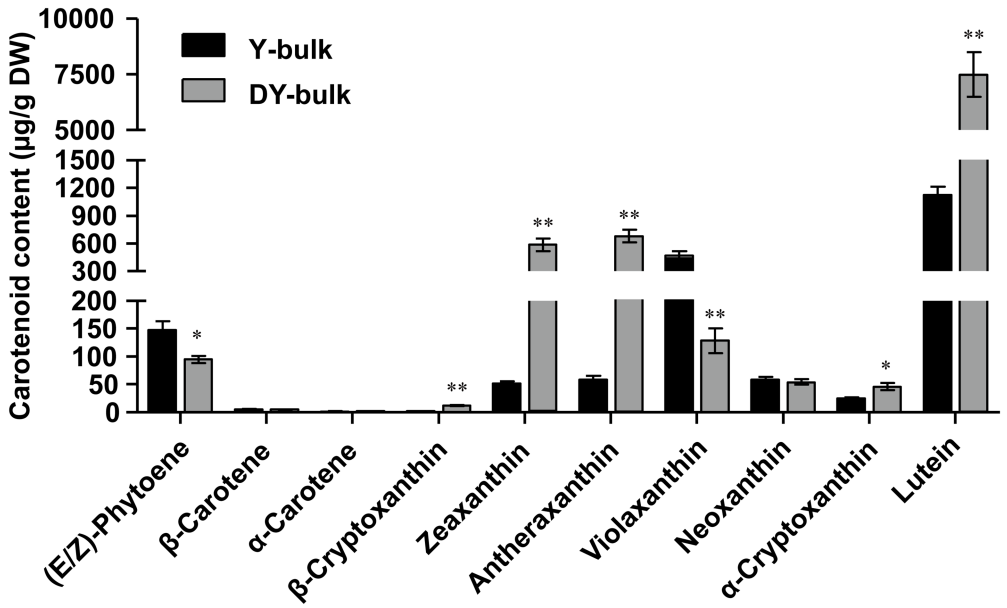


FIGURE 2 | Carotenoid composition in petals from Y-bulk and DY-bulk. Carotenoids extracts from yellow and dark yellow petals were subjected to UPLC-APCI-MS/MS system under saponification treatment. Error bars indicate SE ($n = 3$). Value of $*p < 0.05$ (Student's *t*-test). Value of $**p < 0.01$ (Student's *t*-test).

(**Supplementary Figure 1**; **Supplementary Table 6**), which was consistent with the genetic analysis showing the dark yellow petal color trait was controlled by a single recessive nuclear gene.

To validate the BSR-Seq results, 35 KASP markers in the candidate region were developed and used to screen the two parents. The results showed that 13 KASP markers (**Supplementary Table 3**) exhibited good polymorphism. These 13 markers were further genotyped in 94 F_2 plants for linkage analysis (**Supplementary Table 7**). The results showed that the marker SH-K27 co-segregated with the *Br-dyp1* gene in the preliminary mapping population (**Figure 3A**). There was one recombinant individual between *Br-dyp1* and SH-K25 and SH-K29. The genetic distances between the *Br-dyp1* locus and SH-K25 and SH-K29 were 0.6 and 0.5 cM, respectively (**Figure 3A**). The order of the markers in the genetic map is consistent with that in the physical map (**Figure 3A**).

To fine-map the *Br-dyp1* locus, we screened 743 dark-yellow-petaled F_2 plants using the flanking markers SH-K34 and SH-K11 and identified 54 recombinants. All the 54 recombinants were further genotyped using SH-K35, SH-K24, SH-K25, SH-K27, and SH-K29, based on which 11 recombinants (type 4 and type 8) were identified (**Figure 3B**). Then, eight new markers were developed (**Supplementary Table 3**) and were further used to screen all the 11 recombinants using the KASP assay. The results delimited the *Br-dyp1* gene to a 53.6 kb interval between markers SH-K38 and SH-K43, each with two recombinants (type 5 and type 6; **Figure 3B**). Four markers, namely, SH-K41, SH-K36, SH-K37, and SH-K27, co-segregated with the *Br-dyp1* gene in the fine-mapping population (**Figure 3B**).

Candidate Gene Analysis

DNA sequences in the fine-mapping interval (53.6 kb) of *Br-dyp1* were analyzed according to the *B. rapa* reference genome. Totally, 11 annotated or predicted genes were identified in the mapping region (**Table 2**). Among them, *Bra037130* (*BraA09.ZEP*), a homolog of *ZEP* in *Arabidopsis*, could be the candidate gene (**Table 2**). *ZEP* encodes a zeaxanthin epoxidase that catalyzes the conversion of zeaxanthin to antheraxanthin and

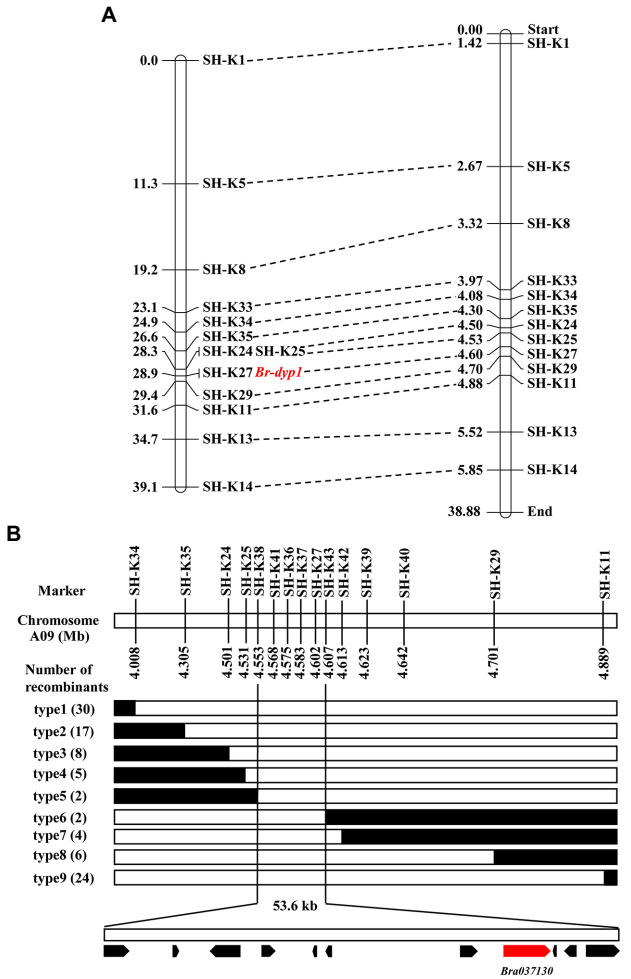


FIGURE 3 | Initial and fine-mapping of the *Br-dyp1* gene in Chinese cabbage. **(A)** Initial mapping of *Br-dyp1*. Genetic map of *Br-dyp1* is on the left, with cM as the unit. The corresponding physical map (right, unit: Mb) are also shown. **(B)** Fine-mapping of *Br-dyp1*. The *Br-dyp1* gene was delimited to an interval between SH-K38 and SH-K43 on chromosome A09, with an estimated physical length of 53.6 kb, and 11 genes were annotated in this region based on the reference genome sequence. The genetic structure of each recombinant type is depicted as white for homozygous dark yellow petal color phenotype, black for heterozygous alleles, respectively. The number of each recombinant type is indicated in the brackets.

violaxanthin in the carotenoid biosynthesis pathway (Liu et al., 2020b).

Next, we examined the expression of the candidate gene *Bra037130* (*BraA09.ZEP*) in different tissues of the two parent lines. qRT-PCR analysis using primer pairs Br-dyp1-qF1 and Br-dyp1-qR1 (**Supplementary Table 8**) revealed that the expression pattern of *Bra037130* was significantly different between the parental lines. In any of the seven tissues that we examined, the expression of *Bra037130* was much lower in the dark-yellow-petaled parent SD369 than in the yellow-petaled parent R16-11 (**Figure 4A**). The highest levels of *Bra037130* (*BraA09.ZEP*) were present in the stamens and petals in both parental lines (**Figure 4A**).

TABLE 2 | Annotated genes in the candidate interval of the *Br-dyp1* locus.

Gene name	Gene position on A09	Arabidopsis homolog	Gene function
Bra037123	4,553,464–4,555,964	*AT3G46730*	NB-ARC domain-containing disease resistance protein
Bra037124	4,560,614–4,561,153	*AT5G67190*	Encodes a member of the DREB subfamily A-5 of ERF/AP2 transcription factor family; DEAR2
Bra037125	4,564,557–4,567,632	*AT5G67170*	SEC-C motif-containing protein
Bra037126	4,569,807–4,571,165	*AT5G67160*	Encodes a member of the BAHD acyltransferase superfamily; EPS1
Bra037127	4,575,174–4,575,557	*AT5G67070*	Rapid Alkalinization Factor; RALF34
Bra037128	4,576,524–4,577,141	*AT5G67060*	Encodes a bHLH transcription factor; HEC1
Bra037129	4,590,446–4,592,182	*AT5G67050*	alpha/beta-Hydrolases superfamily protein
Bra037130	4,594,979–4,599,768	*AT5G67030*	Zeaxanthin epoxidase; ZEP
Bra037131	4,600,167–4,600,430	*AT5G66985*	Hypothetical protein; HUP44
Bra037132	4,601,342–4,602,476	*AT5G66980*	AP2/B3-like transcriptional factor family protein
Bra037133	4,603,501–4,606,981	*AT5G66960*	Prolyl oligopeptidase family protein

To characterize the sequence of the candidate genes in the parental lines, the primer pair Br-dyp1-F and Br-dyp1-R2 (**Supplementary Table 8**) were designed. Sequence analysis indicated that the candidate gene of R16-11 was 3,020 bp in length and contained 14 exons and 13 introns (**Figure 5A**; **Supplementary Figure 2**). The CDS of the candidate gene in R16-11 was 1965 bp in length (**Supplementary Figure 2**). Sequence alignment showed that there were 43 SNP variations and 11 InDel variations between the genomic sequences of R16-11 and SD369 (**Supplementary Table 9**; **Supplementary Figure 3**). Among these variations, the most significant sequence variation was a 679 bp insertion located at 240 bp of the gDNA, within the first exon in SD369 (**Figure 5A**; **Supplementary Table 10**). The 679 bp insertion caused a premature stop codon at the 93 a.a position (**Figure 5B**). Amino acid sequence alignment indicated that the deduced amino acid sequence of *Bra037130* (*BraA09.ZEP*) in R16-11 was highly identical to the ZEP protein sequence in *Arabidopsis* (**Supplementary Figure 4**), and it contained four conserved motifs: two lipocalin conserved motifs (145–162 a.a and 264–283 a.a), a long monooxygenase domain (223–429 a.a), and a Forkhead-associated domain (579–625 a.a; **Supplementary Figure 4**). The 679 bp insertion in *BraA09.ZEP* caused the loss of the four conserved domains

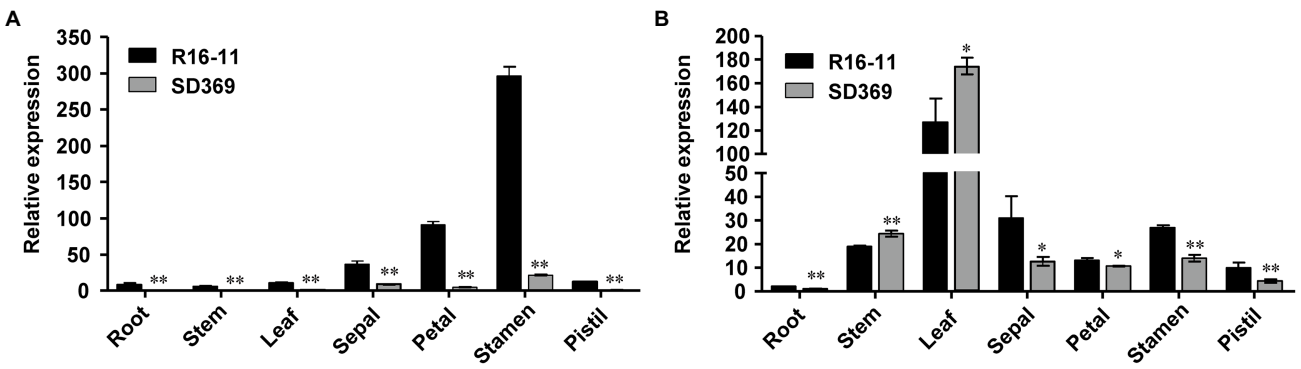


FIGURE 4 | Gene expression data analysis. Quantitative RT-PCR of *Br-dyp1* (*Bra037130*; **A**) and *BraA07.ZEP* (*Bra012127*; **B**) in different tissues of the two parents. The *BrGAPDH* were used as an internal control. Error bars indicate SE ($n = 3$). Value of $*p < 0.05$ (Student's *t*-test). Value of $**p < 0.01$ (Student's *t*-test).

and ultimately resulted in the loss of function of the ZEP protein in dark yellow petals.

Based on the 679 bp insertion in dark-yellow-petaled parent SD369, a functional marker Br-dyp1-InDel (primers pair Br-dyp1-ful-F and Br-dyp1-sp-R1; **Supplementary Table 8**), which could amplify a 1,644-bp and 965-bp product from line SD369 and R16-11, respectively, were developed and were assayed in different materials. The results showed that Br-dyp1-InDel co-segregated with the petal color phenotype in the F_2 population (**Figure 5C**). Furthermore, eight yellow-petaled materials, eight dark yellow-petaled materials, eight orange-petaled materials, and eight white-petaled materials were genotyped using Br-dyp1-InDel. As expected, all the eight dark yellow-petaled materials showed the same genotype as SD369, and all eight yellow-petaled materials showed the same genotype as R16-11 (**Figure 5D**). Interestingly, the eight orange-petaled and eight white-petaled materials also exhibited the same genotype as R16-11 (**Figure 5E**), which implied that the genes controlling the orange and white petal color trait were different from *BraA09.ZEP* and the 679 bp insertion only existed in dark-yellow-petaled materials. Overall, these findings suggest that the *Bra037130* (*BraA09.ZEP*) gene is the most promising candidate gene for the dark yellow petal color gene *Br-dyp1* in Chinese cabbage.

The Coding Sequence and Expression Pattern of *BraA07.ZEP* Show Only a Little Difference Between Dark Yellow and Yellow Petals

Given that the coding sequence of *BraA07.ZEP* (*Bra012127*) was very similar to that of *Br-dyp1* (*BraA09.ZEP*), with 87.02% identity, we designed a primer pair, ZEP-A07-ful-F and ZEP-A07-ful-R (**Supplementary Table 8**), to amplify the full-length CDS of *BraA07.ZEP* in R16-11 and SD369. The CDS of *BraA07.ZEP* from R16-11 and SD369 were submitted to GenBank under the accession numbers OL436222 (R16-11) and OL436223 (SD369). Sequence alignment revealed that there were 29 SNP variations between the coding sequences of R16-11 and SD369 (**Supplementary Table 10**; **Supplementary Figure 5**). Among the 29 SNPs, 27 SNPs were synonymous mutations, and only 2 SNPs caused non-synonymous mutations, which did not affect the protein function (**Supplementary Table 10**; **Supplementary Figure 6**). The expression pattern of *BraA07.ZEP* in the parental lines was also checked using primer pair ZEP-A07-qF and ZEP-A07-qR. The results revealed that the highest transcript levels were detected in the leaves, whereas low levels were found in petals and other tissues (**Figure 4B**). Furthermore, the expression of *BraA07.ZEP* in the petals of R16-11 was only slightly higher than that in SD369 (**Figure 4B**), which indicated a functional divergence between *BraA07.ZEP* and *Br-dyp1* (*BraA09.ZEP*).

Transcriptome Analysis in Dark Yellow and Yellow Petals

To examine the global effect of the *Br-dyp1* (*BraA09.ZEP*) mutation on gene expression in Chinese cabbage, RNA-seq analysis was employed to profile gene expression differences in petals from the Y-bulk and DY-bulk, each with three replicates. Approximately 290 million clean reads were generated for the six samples and 83.9–87.5% were uniquely mapped to the *B. rapa* (Chiifu-401) reference genome (**Supplementary Table 5**). All the clean reads were deposited in the NCBI Short Read Archive database under accession number PRJNA779176. Statistical analysis identified 835 DEGs with at least two-fold changes between the Y-bulk and DY-bulk from the three biological replicates (FDR $\leq$ 0.05). Among these DEGs, 248 genes were upregulated and 587 genes were downregulated in the dark yellow petals.

GO enrichment analysis of the 587 downregulated genes revealed that most DEGs were assigned to the "chloroplast stroma (GO: 0009570)" and "chloroplast thylakoid membrane (GO: 0009535)" terms in the cellular component category (**Figure 6A**), which was compatible with the ZEP localization in chloroplasts (Rock et al., 1992). Zeaxanthin is involved in non-photochemical quenching (NPQ) and thylakoid stacking, and thus affects the PSII function (Rock et al., 1992). Additionally, zeaxanthin serves important functions as an antioxidant in the lipid phase of the membrane and is likely to act as a key

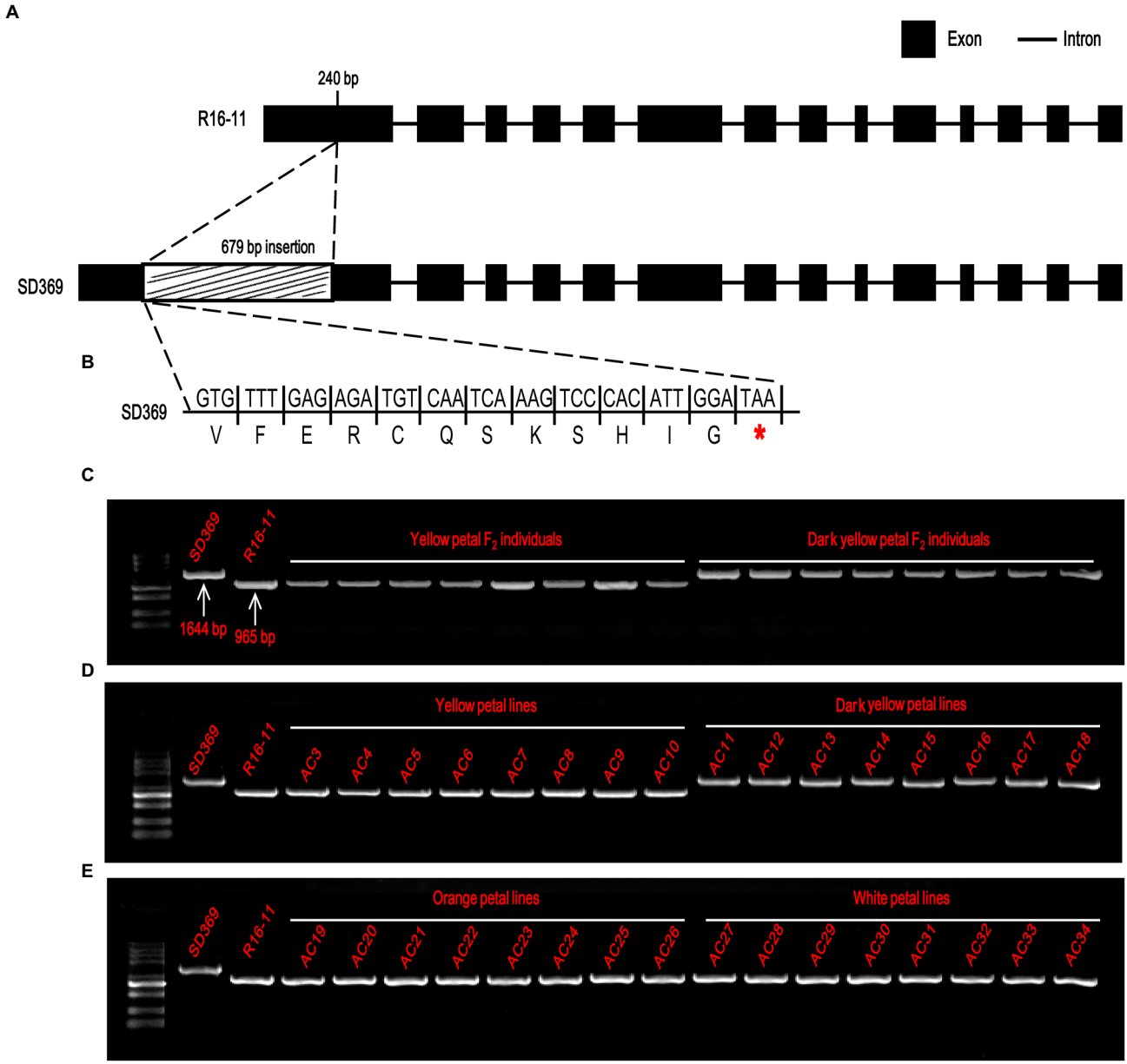


FIGURE 5 | Candidate gene analysis of *Br-dyp1*. **(A)** *Br-dyp1* includes 14 exons and 13 introns in the two parents. **(B)** A 679-bp insertion in dark-yellow-flowered SD369 caused a premature stop codon. **(C–E)** Validation of the functional marker Br-dyp1-InDel in F_2 individuals **(C)**, in 8 yellow and 8 dark-yellow-flowered materials **(D)**, and in eight orange and eight white flowered materials **(E)**.

component in the memory of the chloroplast with respect to preceding photo-oxidative stress (Jahns and Holzwarth, 2012). In this study, genes participating in the "response to heat (GO: 0009408)," the "response to hydrogen peroxide (GO: 0042542)," the "fructose 1,6-bisphosphate metabolic process (GO: 0030388)," and "carbohydrate transport (GO: 0008643)" were significantly enriched in the biological process category (**Figure 6A**), which implied that due to the increased accumulation of zeaxanthin in dark yellow petals, the photo and heat stress and reactive oxygen were removed, so the genes involved in heat and hydrogen peroxide responses were downregulated. Furthermore, due to the sustained energy dissipation by zeaxanthin, genes with function of "UDP-glycosyltransferase activity (GO: 0008194)" and "sugar transmembrane transporter activity (GO: 0051119)" were downregulated (**Figure 6A**). Interestingly, GO enrichment analysis of the 248 upregulated genes showed that "cell wall biogenesis (GO: 0042546)," "cell wall organization (GO: 0071555), and "xyloglucan metabolic process (GO: 0010411)" were the top three significantly enriched terms in the biological process

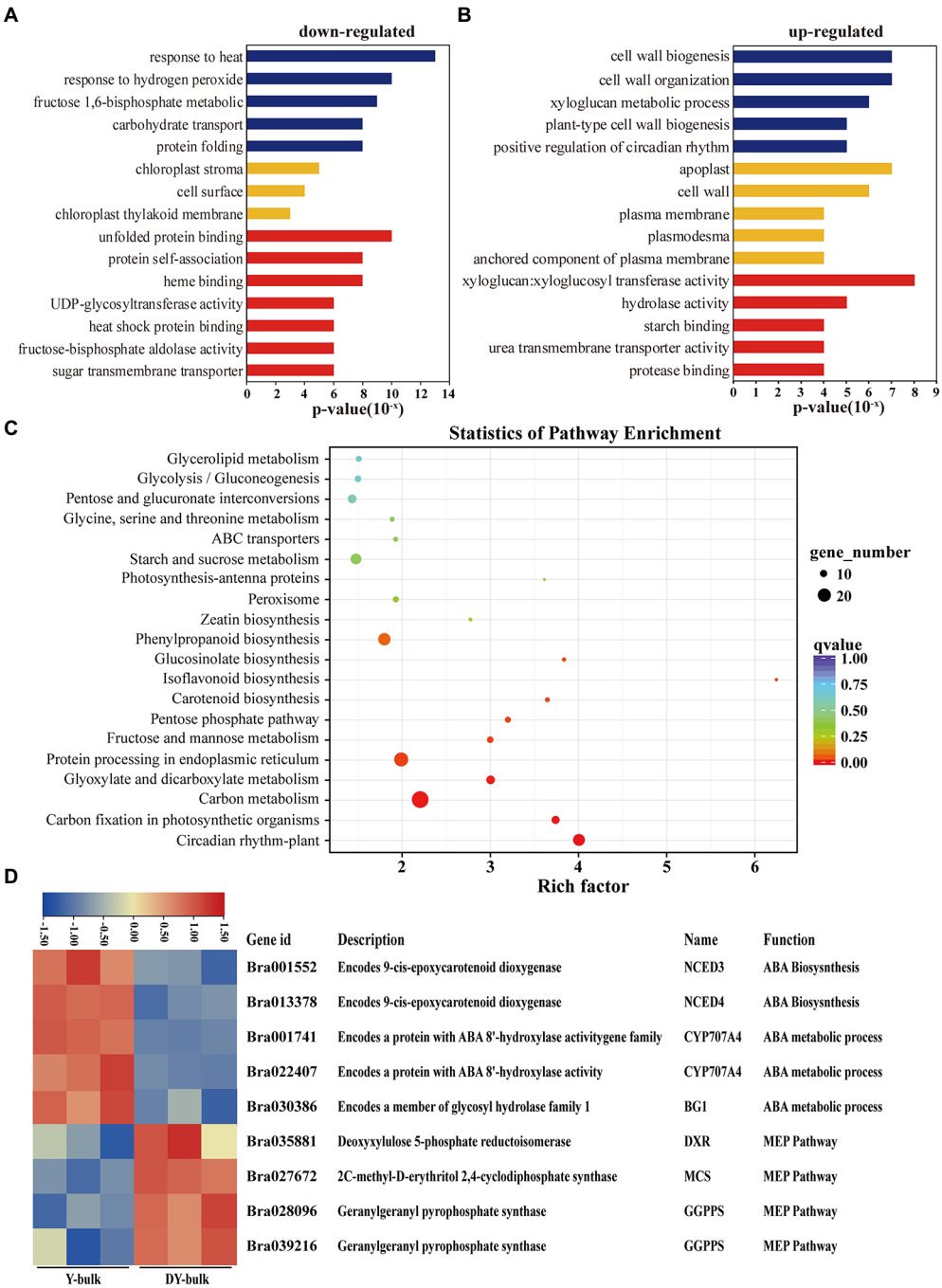


FIGURE 6 | Transcriptome analysis in yellow and dark yellow petals. **(A,B)** GO terms that were significantly enriched in 587 downregulated genes **(A)** and in 278 upregulated genes in the DY-bulk **(B)**. **(C)** Scatter plot of top 20 enriched KEGG pathways. Rich factor is the ratio of the DEG number to the background number in a certain pathway. The size of the dots represents the number of genes, and the color of the dots represents the range of the *q*-value. **(D)** Differentially expressed genes related to the carotenoid metabolism and flux. The heatmap colors are shown in log2 (FPKM). Three biological replicates of the W-bulk and G-bulk are shown.

category, and large proportions of DEGs were assigned to the "apoplast (GO: 0048046)" and "cell wall (GO: 0005618)" in the cellular component category (**Figure 6B**). Xyloglucan endotransglucosylase/hydrolase genes (*XTHs*), which encode proteins with xyloglucan:xyloglucosyl transferase activity (GO: 0016762; **Figure 6B**), are involved in petal abscission in rose (Singh et al., 2011). In our study, the upregulated expression of *XTHs* in dark yellow petals might have caused faster petal abscission to save energy and to compensate sustained energy dissipation by zeaxanthin.

KEGG pathway enrichment analysis revealed that "carotenoid biosynthesis" was one of the most significantly enriched pathways (**Figure 6C**). When focusing on specific genes, in addition to the downregulation of the candidate gene *Bra037130* (*BraA09.ZEP*) in the DY-bulk, five genes involved in carotenoid degradation and ABA biosynthetic and metabolic processes, including *NECD3* (*Bra001552*), *NCED4* (*Bra013378*), *CYP707A4* (*Bra001741* and *Bra022407*) and *BG1* (*Bra030386*; **Figure 6D**), were also found to be significantly downregulated (FDR ≤ 0.05 and |log2 (fold change)| ≥ 1). Notably, most genes involved in carotenoid biosynthesis, such as *PSY* (*Bra006391*), *PDS* (*Bra010751*), *ZDS* (*Bra040411*), *LCYB* (*Bra022892* and *Bra029825*), and *LCYE* (*Bra006838*), were slightly downregulated (FDR ≤ 0.05 and |log2 (fold change)| < 1; **Supplementary Table 11**). However, some of the genes in the MEP pathway, such as *DXR* (*Bra035881*), *MCS* (*Bra027672*), and *GGPPS* (*Bra028096* and *Bra039216*; **Supplementary Table 11**), were slightly upregulated, which might partially contribute to the increased total carotenoid levels. In recent years, increasing evidence of cross-talk between flavonoid and carotenoid pathways has been obtained (Davuluri et al., 2005; Meng et al., 2019; Liu et al., 2020b). In this study, most genes involved in isoflavonoids and flavonoids biosynthesis were downregulated (**Supplementary Table 11**), which was consistent with the previous study (Liu et al., 2020b). The "circadian rhythm" pathway was significantly enriched, and most of the genes in this pathway were downregulated (**Supplementary Table 11**), which concurred that the transcription of ZEP gene follows a diurnal rhythm (Audran et al., 1998; Thompson et al., 2000). Furthermore, genes in the "carbon fixation in photosynthetic organisms" pathway were significantly enriched, and most of these genes were downregulated (**Supplementary Table 11**), which was consistent with the results of GO enrichment analysis, suggesting that the loss-of-function of ZEP caused increased accumulation of zeaxanthin and disturbed antenna assembly and affected the photosynthesis system.

DISCUSSION

The present study successfully fine-mapped the *Br-dyp1* gene to a physical interval of 53.6 kb. Many lines of evidence revealed that the *Bra037130* (*BraA09.ZEP*) gene is the candidate gene for *Br-dyp1*. First, the functional annotation analysis of 11 genes within the 53.6 kb interval revealed that only one gene, *Bra037130*, homologous to *ZEP* in *Arabidopsis*, was involved in carotenoid biosynthesis. Second, the expression level of *Bra037130* (*BraA09.ZEP*) in flower tissues was much lower in deep yellow petals than in yellow petals. Third, sequence alignment showed that the 679 bp insertion in dark yellow petals caused a premature stop codon, thus causing the loss of function of the ZEP enzyme. Most importantly, we developed a functional marker for the candidate gene, and validation showed that this functional marker co-segregated with the petal color phenotype. Furthermore, carotenoid profile analysis showed increased accumulation of zeaxanthin and a reduction of violaxanthin, which concurred with the ZEP function (Marin et al., 1996; Gonzalez-Jorge et al., 2016). Above all, a 679 bp insertion of *Bra037130* (*BraA09.ZEP*) in SD369 was the main reason that caused the dark yellow petal color phenotype. The developed functional marker can be used for molecular-assisted breeding and for developing new ornamental varieties with visual appeal, which has profound significance.

The gene structure annotation of *Bra037130* (*BraA09.ZEP*) in *B. rapa* V1.5 reference genome has some mistake. When we cloned the candidate gene *Bra037130* (*BraA09.ZEP*) for *Br-dyp1*, we first designed the primer pair Br-dyp1-F and Br-dyp1-R1 (**Supplementary Table 8**) to amplify the full-length sequence according to the *B. rapa* V1.5 reference genome. The results showed that no amplification products could be detected using cDNA from SD369 or R16-11 as a template. Another fragment-amplifying primer pair Br-dyp1-qF2 and Br-dyp1-qR2 (**Supplementary Table 8**), which targeted the eighteenth exon of *Bra037130* according to the *B. rapa* V1.5 annotation, were designed and still no amplification products were detected using cDNA from either parent as the templates. However, the qRT-PCR results (**Figure 4A**) and the RNA-seq analysis (**Figure 5D**) showed that the candidate gene *Bra037130* did express in the parental lines. Thus, there must be some mistake in the *Bra037130* annotation in *B. rapa* V1.5. We subjected the CDS of *Bra037130* to a BLAST search against the *B. rapa* V3.0 reference genome and found that the gene corresponding to *Bra037130* in *B. rapa* V3.0 was *BraA09g009220.3C*. The *BraA09g009220.3C* gene owned 14 exons and 13 introns, whereas *Bra037130* had 18 exons and 17 introns (**Supplementary Table 12**), indicating a considerable difference between the *B. rapa* V1.5 and V3.0 annotations. Therefore, we designed another full-length primer pair BrZEP-ful-F and BrZEP-ful-R2 to amplify the full-length sequence according to the *B. rapa* V3.0 reference genome, and both the full-length gDNA and cDNA could be amplified in the two parent lines (**Supplementary Figure 2**). Thus, the gene structure annotation of *Bra037130* in *B. rapa* V3.0 reference genome is corrected according to our experiments results.

Through LC-APCI-MS/MS analysis with saponification, we observed that lutein was the most abundant carotenoid in yellow petals, whereas violaxanthin was the second most abundant carotenoid, accounting for approximately 57.7 and 24.1% of the total carotenoids in the Y-bulk, respectively. These results differed from our previous study (Yang et al., 2021) and another study conducted in Chinese cabbage (Zhang et al., 2020), in which the violaxanthin was the most abundant carotenoid and lutein was the second most abundant carotenoid. It has been reported that deepoxidation to zeaxanthin is favored in high-light conditions, while epoxidation to violaxanthin predominates under moderate-light conditions (Kalituho et al., 2007; Gonzalez-Jorge et al., 2016). The reasons for the above difference might due to the different seasons and different places for the petal sample collection. Because petals were collected in November and in greenhouse in our previous study (Yang et al., 2021), in which the sunlight was mild and the epoxidation to violaxanthin was favored, whereas we collected the petals in June and in an open field for this study, where the sunlight was bright and strong, so the de-epoxidation to zeaxanthin dominated.

A common feature of *zep* mutant leaves is the increased accumulation of zeaxanthin and decreased production of antheraxanthin and violaxanthin (Rock and Zeevaart, 1991; Marin et al., 1996; Niyogi et al., 1998; Liu et al., 2020b). As expected, in our study, the amount of zeaxanthin increased about 11.2-fold in DY-bulk than that in Y-bulk, and violaxanthin decreased about 3.6-fold in DY-bulk. Unexpectedly, the antheraxanthin increased about 11.5-fold in DY-bulk, which was different from the findings of previous studies (Rock and Zeevaart, 1991; Marin et al., 1996; Gonzalez-Jorge et al., 2016). We speculated that there were at least two mechanisms that might explain this difference. First, the enzyme zeaxanthin epoxidase might show substrate specificity. *Phaeodactylum tricornutum* contains three copies of *ZEP*, which exhibit different catalytic activities and substrate specificities (Eilers et al., 2016). In this study, the mutation of *Br-dyp1* (*BraA09.ZEP*) mainly disturbed the epoxidation from antheraxanthin to violaxanthin, and the enzyme encoded by *Br-dyp1* (*BraA09.ZEP*) might therefore exhibit substrate specificity for antheraxanthin. Second, the genome of Chinese cabbage has undergone genome triplication, and another paralogous gene, *BraA07.ZEP*, might contribute to epoxidation from zeaxanthin to antheraxanthin. Zeaxanthin epoxidases is present in only one gene copy in the model plants *Arabidopsis* or rice (Rock and Zeevaart, 1991; Agrawal et al., 2001). However, there were two copies of *ZEP* found in Chinese cabbage, and the sequence identity between *BraA07.ZEP* and our candidate gene, *Br-dyp1* (*BraA09.ZEP*), was high to 87.02% in CDS. *BraA07.ZEP* was mainly expressed in leaves, whereas *Br-dyp1* (*BraA09.ZEP*) was mainly expressed in flower tissues, the tissue-specific expression pattern showed a functional divergence, which was consistent with the results obtained in *B. napus* (Liu et al., 2020b). Although *BraA07.ZEP* was mainly expressed in leaves, its transcripts could also be detected in petals. Hence, we suspected that the enzyme encoded by *BraA07.ZEP* might partially compensate for the loss-of-function mutation of *Br-dyp1* (*BraA09.ZEP*) in dark yellow petals and might be responsible for epoxidation from zeaxanthin to antheraxanthin. A transgenic line of *A. thaliana* with partly disabled zeaxanthin epoxidase activity also showed increased levels of zeaxanthin and antheraxanthin and decreased levels of violaxanthin (Nowicka et al., 2009), which is same with our results, and further confirming our above speculation.

In addition to the increased zeaxanthin and antheraxanthin, other xanthophylls especially the lutein increased about 6.7-fold in DY-bulk, and the total content of carotenoids showed a remarkable 4.7-fold increase in the DY-bulk, which can also be observed in maturing *Arabidopsis* seeds (Gonzalez-Jorge et al., 2016). Two possible reasons might explain this result. First, the enzyme ZEP is not only responsible for the epoxidation from zeaxanthin to violaxanthin, but may also be responsible for the epoxidation from lutein to lutein epoxide. Lutein epoxide is widespread among photosynthetic and non-photosynthetic plant tissues (Garcia-Plazaola et al., 2007), which has been detected in chromoplasts from flowers (Tai and Chen, 2000), fruits (Watanabe and Takahashi, 1999), seeds (Edelenbos et al., 2001), and tubers (Lu et al., 2001). The loss of ZEP activity in dark yellow petals disturbed not only the xanthophyll cycle but also the lutein epoxide cycle, thus causing increases in the accumulation of zeaxanthin and lutein. Second, the mutation of the gene *Br-dyp1* (*BraA09.ZEP*) blocked the carotenoid flux, impaired the carotenoid degradation, and disturbed the flux from carotenoid to ABA biosynthesis, thus causing the bias from the β-branch to the α-branch, so the lutein increased and the total carotenoid increased. Our results were consistent with another study on a *ZEP* mutant, in which the lutein content was increased 2.2-fold and the total seed carotenoids showed a remarkable 6-fold increase relative to the wild type (Gonzalez-Jorge et al., 2016). In *B. napus*, the disruption of the gene *CCD4* impairs the carotenoid degradation and disturbs the carotenoid flux and causes the total carotenoid to increase approximately 42-fold, ultimately changing the flower color from white to yellow (Zhang et al., 2015). Finally, some genes in the MEP pathways were upregulated in the DY-bulk, while some genes for ABA biosynthesis were downregulated, which provided more precursors for carotenoid biosynthesis and alleviated the carotenoid degradation and loss, which might also contribute to the increased total carotenoids in dark yellow petals.

CONCLUSION

The present study delimited the *Br-dyp1* gene responsible for the dark yellow petal color trait in Chinese cabbage. The *Br-dyp1* gene was fine-mapped to an interval of 53.6 kb *via* BSR-Seq and linkage analysis. Through functional annotation, expression profile and sequence variation analysis, *Bra037130* (*BraA09.ZEP*) which encodes a zeaxanthin epoxidase, was the most likely candidate gene for *Br-dyp1*. *BraA09.ZEP* is involved in the epoxidation from zeaxanthin to violaxanthin. A 679 bp insertion in dark yellow petals caused a premature stop codon, and, thus, the loss of ZEP enzyme function, which affected carotenoid metabolism and caused an increase in the accumulation of total carotenoids. Moreover, we developed and validated the functional marker Br-dyp1-InDel for *Br-dyp1*. This achievement is an important advance for molecular research on flower pigmentation in Chinese cabbage.

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

X-WZ and YY conceptualized the experiments and provided the funding resource. SY drafted the manuscript. HL and YZ performed the experiments and analyzed the data. XW, HS,

ZW, and XZ participated in drafting the article and revising it critically. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by Zhongyuan Scholar Program (202101510003), the China Agriculture Research System (CARS-23-G-15), Sci-Tech Innovation Team of Henan Academy of Agricultural Sciences (2021TD06), and the Self-dependent Innovation Program in Henan Academy of Agricultural Science (2121ZC23).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.841328/full#supplementary-material

REFERENCES

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095

Agrawal, G. K., Yamazaki, M., Kobayashi, M., Hirochika, R., Miyao, A., and Hirochika, H. (2001). Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion. Tagging of a zeaxanthin epoxidase gene and a novel *ostatc* gene. *Plant Physiol.* 125, 1248–1257. doi: 10.1104/pp.125.3.1248

Alboresi, A., Dall'osto, L., Aprile, A., Carillo, P., Roncaglia, E., Cattivelli, L., et al. (2011). Reactive oxygen species and transcript analysis upon excess light treatment in wild-type *Arabidopsis thaliana* vs a photosensitive mutant lacking zeaxanthin and lutein. *BMC Plant Biol.* 11:62. doi: 10.1186/1471-2229-11-62

Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., et al. (2012). The path from β-carotene to carlactone, a strigolactone-like plant hormone. *Science* 335, 1348–1351. doi: 10.1126/science.1218094

Alexa, A., and Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. R package version 2.18.0.

Audran, C., Borel, C., Frey, A., Sotta, B., Meyer, C., Simonneau, T., et al. (1998). Expression studies of the zeaxanthin epoxidase gene in nicotiana plumbaginifolia. *Plant Physiol.* 118, 1021–1028. doi: 10.1104/pp.118.3.1021

Davuluri, G. R., van Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., et al. (2005). Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes. *Nat. Biotechnol.* 23, 890–895. doi: 10.1038/nbt1108

Edelenbos, M., Christensen, L. P., and Grevsen, K. (2001). HPLC determination of chlorophyll and carotenoid pigments in processed green pea cultivars (*Pisum sativum* L.). *J. Agric. Food Chem.* 49, 4768–4774. doi: 10.1021/jf010569z

Eilers, U., Dietzel, L., Breitenbach, J., Buchel, C., and Sandmann, G. (2016). Identification of genes coding for functional zeaxanthin epoxidases in the diatom *Phaeodactylum tricornutum*. *J. Plant Physiol.* 192, 64–70. doi: 10.1016/j.jplph.2016.01.006

Frey, A., Boutin, J. P., Sotta, B., Mercier, R., and Marion-Poll, A. (2006). Regulation of carotenoid and ABA accumulation during the development and germination of *Nicotiana plumbaginifolia* seeds. *Planta* 224, 622–632. doi: 10.1007/s00425-006-0231-2

Gao, Z., Liu, Q., Zheng, B., and Chen, Y. (2013). Molecular characterization and primary functional analysis of PeVDE, a violaxanthin de-epoxidase gene from bamboo (*Phyllostachys edulis*). *Plant Cell Rep.* 32, 1381–1391. doi: 10.1007/s00299-013-1450-1

Garcia-Plazaola, J. I., Matsubara, S., and Osmond, C. B. (2007). The lutein epoxide cycle in higher plants: its relationships to other xanthophyll cycles and possible functions. *Funct. Plant Biol.* 34, 759–773. doi: 10.1071/FP07095

Gonzalez-Jorge, S., Mehrshahi, P., Magallanes-Lundback, M., Lipka, A. E., Angelovici, R., Gore, M. A., et al. (2016). ZEAXANTHIN EPOXIDASE activity potentiates carotenoid degradation in maturing seed. *Plant Physiol.* 171, 1837–1851. doi: 10.1104/pp.16.00604

Han, F., Cui, H., Zhang, B., Liu, X., Yang, L., Zhuang, M., et al. (2019). Map-based cloning and characterization of *BoCCD4*, a gene responsible for white/yellow petal color in *B. oleracea*. *BMC Genomics* 20:242. doi: 10.1186/s12864-019-5596-2

Jahns, P., and Holzwarth, A. R. (2012). The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochim. Biophys. Acta* 1817, 182–193. doi: 10.1016/j.bbabio.2011.04.012

Kalituho, L., Rech, J., and Jahns, P. (2007). The roles of specific xanthophylls in light utilization. *Planta* 225, 423–439. doi: 10.1007/s00425-006-0356-3

Kevan, P. G., and Baker, H. G. (1983). Insects as flower visitors and pollinators. *Annu. Rev. Entomol.* 28, 407–453. doi: 10.1146/annurev.en.28.010183.002203

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. (1943). The estimation of map distances from recombination values. *Ann. Eugenics* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

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, L., Yuan, H., Zeng, Y., and Xu, Q. (2016). Plastids and carotenoid accumulation. *Subcell. Biochem.* 79, 273–293. doi: 10.1007/978-3-319-39126-7_10

Liu, Y., Lv, J., Liu, Z., Wang, J., Yang, B., Chen, W., et al. (2020a). Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (*Capsicum annuum* L.). *Food Chem.* 306:125629. doi: 10.1016/j.foodchem.2019.125629

Liu, Y., Ye, S., Yuan, G., Ma, X., Heng, S., Yi, B., et al. (2020b). Gene silencing of *BnaA09.ZEP* and *BnaC09.ZEP* confers orange color in *Brassica napus* flowers. *Plant J.* 104, 932–949. doi: 10.1111/tpj.14970

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

Lou, Y., Sun, H., Li, L., Zhao, H., and Gao, Z. (2017). Characterization and primary functional analysis of a bamboo *ZEP* gene from *phyllostachys edulis*. *DNA Cell Biol.* 36, 747–758. doi: 10.1089/dna.2017.3705

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/s13059-014-0550-8

Lu, W., Haynes, K., Wiley, E., and Clevidence, B. (2001). Carotenoid content and color in diploid potatoes. *J. Am. Soc. Hortic. Sci.* 126, 722–726. doi: 10.21273/jashs.126.6.722

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., et al. (1996). Molecular identification of zeaxanthin epoxidase of Nicotiana plumbaginifolia, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of Arabidopsis thaliana. *EMBO J.* 15, 2331–2342. doi: 10.1002/j.1460-2075.1996.tb00589.x

Meng, Y., Wang, Z., Wang, Y., Wang, C., Zhu, B., Liu, H., et al. (2019). The MYB activator WHITE PETAL1 associates with MtTT8 and MtWD40-1 to regulate carotenoid-derived flower pigmentation in *Medicago truncatula*. *Plant Cell* 31, 2751–2767. doi: 10.1105/tpc.19.00480

Niyogi, K. K., Grossman, A. R., and Bjorkman, O. (1998). *Arabidopsis mutants* define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. *Plant Cell* 10, 1121–1134. doi: 10.2307/3870716

Nowicka, B., Strzalka, W., and Strzalka, K. (2009). New transgenic line of *Arabidopsis thaliana* with partly disabled zeaxanthin epoxidase activity displays changed carotenoid composition, xanthophyll cycle activity and non-photochemical quenching kinetics. *J. Plant Physiol.* 166, 1045–1056. doi: 10.1016/j.jplph.2008.12.010

Park, H. Y., Seok, H. Y., Park, B. K., Kim, S. H., Goh, C. H., Lee, B. H., et al. (2008). Overexpression of *Arabidopsis ZEP* enhances tolerance to osmotic stress. *Biochem. Biophys. Res. Commun.* 375, 80–85. doi: 10.1016/j.bbrc.2008.07.128

Qi, J., Yu, S., Zhang, F., Shen, X., Zhao, X., Yu, Y., et al. (2010). Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Mol. Biol. Report.* 28, 597–604. doi: 10.1007/s11105-010-0185-1

Rock, C. D., Bowlby, N. R., Hoffmann-Benning, S., and Zeevaart, J. A. (1992). The *aba mutant* of *Arabidopsis thaliana* (L.) heynh. Has reduced chlorophyll fluorescence yields and reduced thylakoid stacking. *Plant Physiol.* 100, 1796–1801. doi: 10.1104/pp.100.4.1796

Rock, C. D., and Zeevaart, J. A. (1991). The *aba mutant* of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7496–7499. doi: 10.1073/pnas.88.17.7496

Singh, A. P., Tripathi, S. K., Nath, P., and Sane, A. P. (2011). Petal abscission in rose is associated with the differential expression of two ethylene-responsive xyloglucan endotransglucosylase/hydrolase genes, *RbXTH1* and *RbXTH2*. *J. Exp. Bot.* 62, 5091–5103. doi: 10.1093/jxb/err209

Su, T., Yu, S., Zhang, J. W. F., Yu, Y., Zhang, D., Zhao, X., et al. (2014). Loss of function of the carotenoid Isomerase gene *BrCRTISO* confers Orange color to the inner leaves of Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Mol. Biol. Report.* 33, 648–659. doi: 10.1007/s11105-014-0779-0

Tai, C. Y., and Chen, B. H. (2000). Analysis and stability of carotenoids in the flowers of daylily (*Hemerocallis disticha*) as affected by various treatments. *J. Agric. Food Chem.* 48, 5962–5968. doi: 10.1021/jf000956t

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Thompson, A. J., Jackson, A. C., Parker, R. A., Morpeth, D. R., Burbidge, A., and Taylor, I. B. (2000). Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-*cis*-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid. *Plant Mol. Biol.* 42, 833–845. doi: 10.1023/a:1006448428401

Van Ooijen, J.W. (2006). *JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations of Diploid Species*. Wageningen, Netherlands: Plant Research International BV and Kayazma BV.

Walter, M. H., and Strack, D. (2011). Carotenoids and their cleavage products: biosynthesis and functions. *Nat. Prod. Rep.* 28, 663–692. doi: 10.1039/c0np00036a

Wang, H., Huang, Y., Xiao, Q., Huang, X., Li, C., Gao, X., et al. (2020). Carotenoids modulate kernel texture in maize by influencing amyloplast envelope integrity. *Nat. Commun.* 11:5346. doi: 10.1038/s41467-020-19196-9

Watanabe, K., and Takahashi, B. (1999). Chlorophyll and carotenoid pigments in green- and yellow-fleshed kiwifruit during fruit development and storage. *J. Japan. Soc. Hortic. sci.* 68, 1038–1043. doi: 10.2503/jjshs.68.1038

Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 34, W720–W724. doi: 10.1093/nar/gkl167

Xiong, L., Lee, H., Ishitani, M., and Zhu, J. K. (2002). Regulation of osmotic stress-responsive gene expression by the *LOS6/ABA1* locus in *Arabidopsis*. *J. Biol. Chem.* 277, 8588–8596. doi: 10.1074/jbc.M109275200

Yang, S., Tian, X., Wang, Z., Wei, X., Zhao, Y., Su, H., et al. (2021). Fine mapping and candidate gene identification of a white flower gene *BrWF3* in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). Front. *Plant Sci.* 12:646222. doi: 10.3389/fpls.2021.646222

Yang, S., Yu, W., Wei, X., Wang, Z., Zhao, Y., Zhao, X., et al. (2020). An extended KASP-SNP resource for molecular breeding in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *PLoS One* 15:e0240042. doi: 10.1371/journal.pone.0240042

Zhang, N., Chen, L., Ma, S., Wang, R., He, Q., Tian, M., et al. (2020). Fine mapping and candidate gene analysis of the white flower gene *Brwf* in Chinese cabbage (*Brassica rapa* L.). *Sci. Rep.* 10:6080. doi: 10.1038/s41598-020-63165-7

Zhang, X., Li, R., Chen, L., Niu, S., Chen, L., Gao, J., et al. (2018a). Fine-mapping and candidate gene analysis of the *Brassica juncea* white-flowered mutant *Bjpc2* using the whole-genome resequencing. *Mol. Gen. Genomics.* 293, 359–370. doi: 10.1007/s00438-017-1390-5

Zhang, X., Li, R., Chen, L., Niu, S., Li, Q., Xu, K., et al. (2018b). Inheritance and gene mapping of the white flower trait in *Brassica juncea*. *Mol. Breed.* 38:20. doi: 10.1007/s11032-017-0771-0

Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., et al. (2015). Disruption of a *CAROTENOID CLEAVAGE DIOXYGENASE* 4 gene converts flower colour from white to yellow in brassica species. *New Phytol.* 206, 1513–1526. doi: 10.1111/nph.13335

Zhou, W., Niu, Y., Ding, X., Zhao, S., Li, Y., Fan, G., et al. (2020). Analysis of carotenoid content and diversity in apricots (*Prunus armeniaca* L.) grown in China. *Food Chem.* 330:127223. doi: 10.1016/j.foodchem.2020.127223

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: 15 July 2022
doi: 10.3389/fpls.2022.936958

A Chromosome Level Genome Assembly of a Winter Turnip Rape (*Brassica rapa L.*) to Explore the Genetic Basis of Cold Tolerance

Junyan Wu[1,2†], Xin-Dong Xu[3†], Lijun Liu[2], Li Ma[2], Yuanyuan Pu[1], Wangtian Wang[2], Xue-Yang Hua[3], Jia-Ming Song[3], Kede Liu[4], Guangyuan Lu[5], Yan Fang[2], Xuecai Li[1] and Wancang Sun[1,2]*

[1] College of Agronomy, Gansu Agricultural University, Lanzhou, China, [2] State Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou, China, [3] State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, College of Life Science and Technology, Guangxi University, Nanning, China, [4] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China, [5] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China

OPEN ACCESS

Edited by:
Shoupu He,
National Key Laboratory of Cotton Biology, Institute of Cotton Research (CAAS), China

Reviewed by:
Tongkun Liu,
Nanjing Agricultural University, China
Deshuang Zhang,
Beijing Vegetable Research Center, China

***Correspondence:**
Wancang Sun
18293121851@163.com

[†] These authors have contributed equally to this work

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

Received: 05 May 2022
Accepted: 24 June 2022
Published: 15 July 2022

Citation:
Wu J-Y, Xu X-D, Liu L-J, Ma L, Pu Y-Y, Wang W-T, Hua X-Y, Song J-M, Liu K-D, Lu G-Y, Fang Y, Li X-C and Sun W-C (2022) A Chromosome Level Genome Assembly of a Winter Turnip Rape (Brassica rapa L.) to Explore the Genetic Basis of Cold Tolerance. Front. Plant Sci. 13:936958. doi: 10.3389/fpls.2022.936958

Winter rapeseed (*Brassica rapa L.*) is an important overwintering oilseed crop that is widely planted in northwest China and suffers chronic low temperatures in winter. So the cold stress becomes one of the major constraints that limit its production. The currently existing genomes limit the understanding of the cold-tolerant genetic basis of rapeseed. Here we assembled a high-quality long-read genome of *B. rapa* "Longyou-7" cultivar, which has a cold-tolerant phenotype, and constructed a graph-based pan-genome to detect the structural variations within homologs of currently reported cold-tolerant related genes in the "Longyou-7" genome, which provides an additional elucidation of the cold-tolerant genetic basis of "Longyou-7" cultivar and promotes the development of cold-tolerant breeding in *B. rapa*.

Keywords: *Brassica rapa*, graph-based pan-genome, structural variation (SV), cold tolerance, physiological characteristics

BACKGROUND

Brassica rapa is a crop species of nutritional and economic importance. It is cultivated worldwide as oil and vegetable crops. It belongs to the genus *Brassica*, tribe *Brassiceae* of the family *Brassicaceae*. *B. rapa* (AA, 2n = 20) is one of the three diploid ancestors of *B. napus* (AACC, 2n = 38) and *B. juncea* (AABB, 2n = 36). During diversification, *B. rapa* formed different subspecies and morphotypes, including turnips, leafy greens, such as bok choy and Chinese cabbage, and oilseed crops, such as turnip rape and yellow sarsons (Gómez-Campo and Prakash, 1999; Prakash et al., 2011). Many of the *B. rapa* crops are annual, but turnips, some turnip rape cultivars, and some Chinese cabbage are biennial and require vernalization to flower (Zhao et al., 2007). The oleiferous form, namely *B. rapa* ssp. *oleifera*, or turnip rape, is the third most important *Brassica* oilseed crop after *B. napus* and *B. juncea*, and is widely grown in China, Canada, India, and northern Europe (Ramchiary and Lim, 2011). Before the introduction of *B. napus* into China in the 1930's, winter turnip rape was the major Brassica oilseed crop in the provinces of South China and along the Yangtse River (Wang, 2010), but it hardly survives in North China, including the Loess Plateau and the large areas North to the Yellow River due to the prolonged low temperature and dry weather in

winter. In recent years, several winter turnip varieties including "Longyou-6" (LY6) and "Longyou-7" (LY7) with strong freeze tolerance were released and extensively grown in these regions, which greatly enlarged the area of oilseed rape production in China. The varieties are sowed in late August or early September. Old leaves were beginning to turn yellow from October and all leaves withered in late November as the temperature below −10°C and covered the surface of the land, but the shoot apexes keep alive across the winter. New leaves sprout out from the alive shoot apexes next March when the temperature turns warm. The plantation of these super freeze-resistant varieties not only changes the farming system from single cropping to one and a half or double cropping per year but also increases the land surface coverage during winter and prevents soil erosion of naked land in North China by the strong wind.

A high-quality reference genome is a valuable resource for genetic and genomic studies. The *B. rapa* genome was the first to be sequenced among the *Brassica* species (Wang et al., 2011). The multi-national *B. rapa* Genome Sequencing Project (BrGSP) was launched in 2003, which aimed to obtain the genome sequence of Chinese cabbage accession "Chiifu-401-42" using a BAC-by-BAC strategy (Trick et al., 2007). The first released *B. rapa* cv. "Chiifu-401-42" draft reference genome, v1.5, was assembled using a whole-genome shotgun strategy with Illumina short reads (Wang et al., 2011). However, the first *B. rapa* genome assembly (version 1.5) is only about 283.8 Mb, 58.52% of the estimated genome size (485 Mb) (Wang et al., 2011). The *B. rapa* genome v2.0 was *de novo* assembled with an additional 76G Illumina paired-end reads (≈156×) and 6.5G PacBio single-molecule data (≈13×) (Cai et al., 2017). It was updated to the *B. rapa* genome v2.5 after improving the scaffold order (http://brassicadb.org/brad/datasets/pub/Genomes/Brassica_rapa/V2.0/V2.5/). However, due to the relatively recent whole genome triplication, highly repeated sequences, and complicated centromeric regions, the early three versions of reference genomes assembled mainly using short reads are highly fragmented and contain thousands of discrete contigs and a large number of misassemblies. The inaccuracy in assembly and the low contiguity of these draft assemblies have largely hindered their applications in both genomic and genetic studies of *B. rapa* and other related Brassica species. More recently, a significantly improved *B. rapa* draft genome (v3.0) was assembled using single-molecule PacBio sequencing, optical mapping, and chromosome conformation capture technologies (Hi-C) (Zhang et al., 2018). Relative to the previous reference genomes, the v3.0 assembly reached a contig N50 size of 1.45 Mb, representing a ∼30-fold improvement of contiguity.

B. rapa is a highly diverse and widely cultivated crop species worldwide. The "Chiifu-401-42" reference genome only is not sufficient to capture all or even most of the variants and can hardly satisfy the needs of subsequent functional genomics research and molecular breeding of *B. rapa*. Multiple high-quality reference genomes representing different morphotypes and ecotypes are necessary for a better understanding of the genome structure and genetic basis of morphotype and ecotype differentiation in *B. rapa*. Using the long-read sequencing technologies such as Oxford Nanopore Technology (ONT) and Pacific Biosciences (PacBio), 22 *B. rapa* varieties including morphotypes of Chinese cabbage, turnip, oilseed, taicai, mizuna, and pak choi (pak choi, wutacai, caixin) have been sequenced so far and assembled in high-quality in terms of continuity and completeness of repetitive regions (Belser et al., 2018; Cai et al., 2021; Li et al., 2021). Based on 18 *B. rapa* genomes, structural variations (SVs) were identified and an integrated graph-based pan-genome was constructed. Based on the pan-genome, SVs were genotyped in 524 *B. rapa* genomes, and SVs involved in leafy head domestication were identified (Cai et al., 2021).

However, among the 22 accessions sequenced, none has been indicated to be winter turnip rape. These genomes may not be suitable for elucidating candidate genes associated with cold resistance. To obtain a high-quality reference genome for the identification of genes involved in cold resistance, in this study, we assembled a high-quality genome of "LY7" using PacBio HIFI reads and confirmed the whole genome duplication (WGD) event that occurred in the genome of *B. rapa* (Wang et al., 2011). Combining the genomes of "LY7" and the other 22 *B. rapa* accessions, we constructed a graph-based pan-genome. Joint analysis of pan-genome and RNA-seq data identified two genes, *HDG1* and *BrANS3*, which may be associated with cold tolerance. The reference genome of "LY7" and the graph-based genome will be useful for the identification of cold resistance genes through map-based gene cloning and genome-wide association studies in the future.

RESULTS

Morphological Characteristics of *B. rapa* "Longyou-7"

"LY7" and "Lenox" are two winter turnip rape varieties with strong and weak cold tolerance, respectively. Compared to the weak cold-tolerant "Lenox," "LY7" displayed several distinct character traits such as fewer leaves, smaller and shorter leaves, and longer and larger taproot, which finally leads to its smaller dry leaf weight, larger dry taproot weight, and root/shoot ratio than Lenox (**Table 1**; **Supplementary Figure S1**). In addition, LY7 displayed prostrate growth and its shoot apex meristem (SAM) is beneath the ground (**Figure 1**), which protects it from drastic air temperature change. Furthermore, the leaves of LY7 withered earlier than Lenox, which could reduce water loss and also alleviate further damage transduced from leaves to SAM across winter. In two consecutive years of production experiments in six locations in Gansu province, China, Zhangye, Wuwei, Aolan, Lanzhou, Qingyang, and Jiuquan, the overwintering rate of "Longyou-7" variety of rapeseed reached 90.2% to 97.0% percent, and that of "Lenox" variety rapeseed reached 1.4% to 10.3% percent.

Physiological Characteristics of *B. rapa* "Longyou-7"

When subjected to low temperatures, plants produce increased content of reactive oxygen species (ROS). Excess ROS can result in oxidative damage to cellular membranes and other cellular components, which ultimately leads to cell death.

Malondialdehyde (MDA) is a type of ROS generated by the peroxidation of membrane polyunsaturated fatty acids (Esterbauer et al., 1991) and the content of MDA in the plant is often used as a parameter to evaluate the damage to plant cells due to stress. Plant with lower amounts of MDA under low temperature is generally considered more tolerant to cold. To determine whether ROS accumulation is related to the difference in cold tolerance between LY7 and Lenox, we measured the MDA content of plants grown under natural conditions. LY7 and Lenox were sowed in late August and leaves were collected once every month after sowing until December when the temperature went frozen. MDA contents increased in both LY7 and Lenox as the temperature went down from September to December, and the MDA content in Lenox is significantly higher than that in LY7 (**Supplementary Figure S2**), indicating that cell membranes of LY7 were less damaged by ROS generated by low temperature. Soluble protein has been proven to enhance the cold hardiness of plants, and increasing the content of soluble protein can enhance the cells to retain moisture, improving cold resistance capability (Jung et al., 1967). The content of soluble protein increased from September to October, then decreased, and the content did not show a significant difference between LY7 and Lenox.

ROS-scavenging enzymes play crucial roles in ROS homeostasis, and ascorbate peroxidase (APX), peroxidase (POD), superoxide dismutase (SOD) and catalase (CAT) are

TABLE 1 | Statistics of assemblies and annotation in *B. rapa* "Longyou-7" genome.

Figure	Value
Assembly	
Total contigs	158
Total length (bp)	413,277,474
Min scaffold length (bp)	50,948
Max scaffold length (bp)	72,779,390
Average length (bp)	6,168,320.51
Contig N50 length (bp)	10,226,103
Scaffold N50 length (bp)	42,251,188
(G + C)s (%)	37.05
Annotation	
Repeat length (bp)	223
Repeat ratio (%)	54.35
Total gene number	45,844

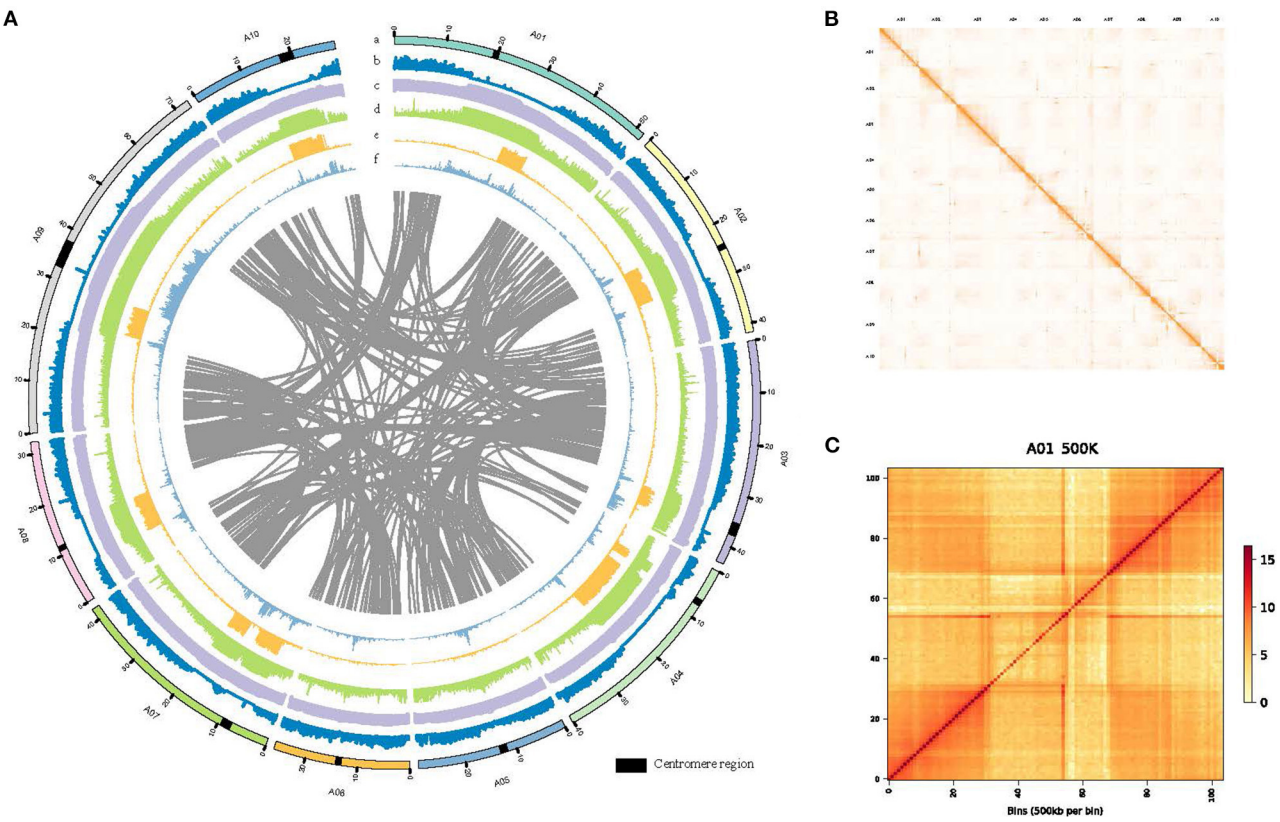


FIGURE 1 | *De novo* genome assembly and annotation of *B.rapa* "Longyou-7" genome. **(A)** Global view of genome features of *B. rapa* "Longyou-7," Track a shows the 10 pseudochromosomes of *B. rapa* "Longyou-7" denoted by different colors. Track b shows the genome-wide distribution of Gene density. Track c shows the density distribution of GC content. Track d displays the density distribution of transposable elements (TE). Track e shows the density distribution of Copia. Track f shows the density distribution of Gypsy. The red triangles represent the location of the centromeres in all chromosomes; **(B)** Genome-wide Hi-C contact matrix of *B. rapa* "Longyou-7" genome; **(C)** Hi-C map of contact between two 500 kb loci in A01 of *B. rapa* "Longyou-7" genome.

four major enzymes that can reduce the accumulation of ROS, weaken the damage to cells and improve the stress resistance of plants (Apel and Hirt, 2004). We measured the activities of these four enzymes and found that the activities of these four enzymes in LY7 are significantly higher than that in Lenox at all time points, suggesting that LY7 could efficiently remove ROS generated by low temperature and protect cells from damage. These physiological indicators may explain the difference in cold tolerance between LY7 and Lenox.

De novo Assembly and Annotation of the "LY7" Genome

To reveal the genetic basis of the super cold tolerance in "LY7," we sequenced and de novo assembled the genome of "LY7" with high fidelity (HiFi) reads generated by PacBio Circular Consensus Sequencing (CCS) technology. A total of 29 G (equivalent to 68× genome coverage) sequencing data, with 1,767,878 HiFi reads with Q30 of 95.83%, was obtained. The HiFi reads have an average read length of 16,440 bp and read length N50 of 16,344 bp. The maximum read length reached 43,710 bp. We also obtained 64 G high-throughput chromosome conformation capture (Hi-C) reads. Assembly with HiFi reads generated a total of 645 contigs, with a contig N50 size of 10.32 Mb, and a total length of 429 Mb. The longest contig reached 41.70Mb (**Table 1**). Scaffolding with Hi-C reads resulted in 67 scaffolds with a scaffold N50 size of 42.25 Mb and a total length of 413 Mb. With the assistance of Hi-C data, we anchored 408 Mb to 10 pseudo-chromosomes, with A09 being the longest (72.77 Mb) and A06 being the shortest (26.18 Mb) (**Table 1**; **Figures 1A,B**). All the 10 chromosomes show continuous Hi-C signals in the heatmap (**Figure 1B**), indicating frequent interactions between adjacent loci, and each chromosome can be seen to occupy a separate territory within the nucleus. Strong intrachromosomal interactions are also observed within two chromosome arms, such as A01, where the centromere region can be seen in the Hi-C heatmap, which has a significantly smaller range of interactions than the two chromosome arms (**Figure 1C**). The GC content of the *B. rapa* genome was 37.05%. A total of 45,844 gene models were annotated. To assess the quality of the genome assembly of "LY7," we evaluated the completeness of gene models using the single copy embryophyte_odb10 BUSCO dataset which contains 1,614 BUSCO gene sets of core conserved plant genes. Of which, 1,595 genes (98.9%) are intact in the "Longyou-7" genome, 5 genes (0.3%) are fragmented, and 14 genes (0.8%) genes are missed in the "LY7" genome, indicating that the genome assembly has good integrity (Seppey et al., 2019). We also assessed the completeness by mapping the RNA-seq reads to the "LY7" genome and found that more than 92% of HiSeq reads could align properly. These results demonstrated the high completeness of the assembled genome (**Supplementary Table S1**).

To analyze repetitive sequences, we searched the genome sequence via a combination approach of de novo structure-based analysis and homology-based comparisons referring to previous methods (Schmutz et al., 2010). More than half (54.35%) of genomic sequences were annotated as repeat elements, which is higher than that in the 16 genomes assembled by Cai et al. (2021). The mean length of repeat elements was 232 bp, which is longer than most other samples analyzed in the previous study (Cai et al., 2021). A total of 976,574 transposable elements (TEs) were identified (**Supplementary Table S2**). As found in other plant genomes, long terminal repeat (LTR)-retrotransposons were the most abundant elements, including 298,537 Copia-like, 33,378 Gypsy-like, and 7,568 unclassified LTR elements, representing 42.5% of all the identified TEs. In addition to class I retrotransposons, 637,091 class II DNA transposons were identified, including 2,496 Tc1/Mariners, 12,784 hATs, 447,564 Mutators, 11,160 PIF/Harbingers and 155,055 Helitrons. These TEs, together with abundant truncated elements and other repetitive fragments, made up 54.81% of the LY7 genome.

The distributions of gene density, GC content, Gypsy, and Copia density on the 10 pseudo-chromosomes are depicted in a circus plot (**Figure 1A**). Gene density in the two arms was much higher than that in the centromeric regions, which could approximately delimit the locations of the centromeres on chromosomes. The centromere sequences were successfully identified for all chromosomes of LY7 (**Figure 1A**). The distribution of GC is contrasting to gene density, with the centromeric regions having much higher GC content than the arms. The distributions of transposons along chromosomes are consistent with the GC content, with the density of transposable elements in the centromeric regions much higher than that in the chromosome arms. The two major transposon superfamilies, Gypsy and Copia, are mainly located in the centromere regions, which are also consistent with the GC content. Whereas the distribution of the Copia superfamily transposons was complementary to that of the Gypsy superfamily transposons in the centromere regions (**Figure 1A**).

Whole Genome Duplication and Evolution

The *B. rapa* genome has experienced a whole-genome triplication (WGT) event relative to the *A. thaliana* genome (Wang et al., 2011), which was reported to play an important role in the speciation and morphotype diversification of *Brassica* plants (Cheng et al., 2014). Synteny dot plot analysis between the "LY7" genome and the *A. thaliana* genome revealed a 3:1 syntenic depth was identified, which also confirms the extra WGT event reported in the previous study (Wang et al., 2011) that *B. rapa* genome underwent (**Figure 2A**). Self-alignment analysis revealed long stretches of duplications within the assembled *B. rapa* "Longyou-7" genome among chromosomes, for example, chromosome 1, 3, 5, and 8 (**Figure 2B**). Intra-chromosomal duplication also exists in *B. rapa* "Longyou-7" genome, which was only found in chromosome 7. Combined with the phylogenetic tree, we can speculate on the occurrence time and species specificity of WGT (**Figure 2C**). Distribution of synonymous substitutions per synonymous site (Ks) of paralogous genes and syntenic blocks for *B. rapa* "Chiifu" (Wang et al., 2011), *B. rapa* "Longyou-7" and *B. oleracea* (Parkin et al., 2014) confirmed two WGT peaks near Ks = 0.32 and Ks = 0.83, and a divergence peak of *B. rapa* with *A. thaliana* near *Ks = 0.43*, the synonymous replacement rate is chosen to be $1.5*10^{-8}$ mutations site/year

refer to the previous study (Song et al., 2020), so the WGT events in *B. rapa* and *A. thaliana* are estimated to have occurred about 10.7 million years ago and 27.7 million years ago, respectively, according to the formula "T = Ks / 2r." Phylogenic analysis was performed with some *Brassica ssp.* related species, the WGT event of *B. rapa* was found to happen after the divergence between *Thlaspi avense* and *Brassica ssp.*, which is at about 19.3 MYA (**Figure 2D**). WGD in plants was reported to have a significant contribution to plant adaption (Wu et al., 2020), which also includes adaptation to ambient temperature. The times of the above events were similar to what had been reported (Song et al., 2020; Cai et al., 2021). On the other hand, the phylogenetic tree of *B. rapa* was constructed using the maximum likelihood method using single-copy ortholog sequences among the "Longyou-7" genome and 18 reported *B. rapa* genomes (Cai et al., 2021), with species *B. oleracea* (Parkin et al., 2014) as the outgroup. The results of the clustering showed that the "Longyou-7" genome differed significantly from the other 18 genomes of *B. rapa* and formed a separate branch so that it was necessary to construct the reference genome of LY7 to deepen the understanding of the genetic diversity of *B. rapa.* (**Figure 2E**).

The abundance of repetitive sequences is thought to be the main challenge of plant genome assembly and long terminal repeat retrotransposons (LTR-RTs) are the dominant repetitive sequences that are poorly assembled in draft genomes. LTR Assembly Index (LAI) is reported to be used for evaluating assembly continuity using LTR-RTs (Ou et al., 2018). Here we calculated the LAI of 10 chromosomes of our *B. rapa* genome. Our research performs an LAI estimate toward the assembled *B. rapa* genome to visualize its assembly quality. LAI score reflexes the assembly quality of repetitive region sequences. The mean LAI score of our genome is 20.04 with a standard deviation is 4.6, which suggests the assembly of our *B. rapa* genome shows high quality. However, some chromosome regions also present low quality (**Figure 3A**). For example, some regions of chromosome A09 are found with LAI scores under 10, which means the assembly of this region might be implausible (Ou

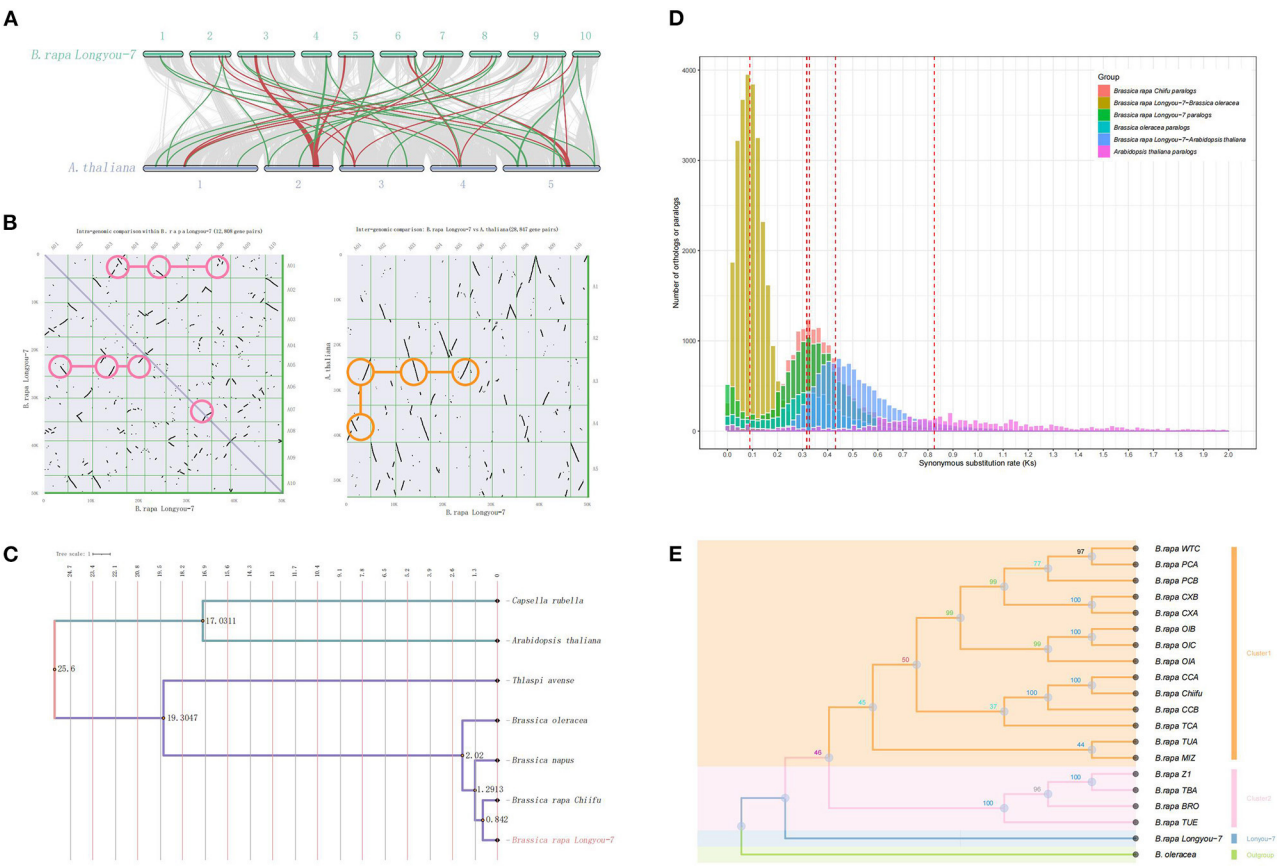


FIGURE 2 | WGT and evolution of *B. rapa* "Longyou-7" genome. **(A)** Macrosynteny patterns reflect a 3:1 syntenic depth between *B. rapa* "Longyou-7" and *A. thaliana* genomes. Green linkages mark examples where three copies of a gene on chromosomes of Arabidopsis are present in the "Longyou-7" genome; **(B)** Dot plot illustrating the comparative analysis within the "Longyou-7" and *Arabidopsis* genomes, the pink circles highlight several major triplication events, the dots represent the synteny gene pairs, and the orange circles highlight an example of 3:1 synteny depth; **(C)** Phylogenetic tree constructed with single-copy orthologs of seven species with divergence time annotated. **(D)** Ks (Synonymous substitution rate) distributions for Brassica rapa paralogs and orthologs with other genomes: *A. thaliana, B. oleracea,* and *B.rapa Chiifu*; **(E)** Species tree constructed using the maximum likelihood method with single-copy ortholog sequences of *B. rapa* "Longyou-7" genome and 18 reported *B. rapa* genomes, with *B. oleracea* as the outgroup.

et al., 2018). The assembly results of the complete repetitive elements provided us with an opportunity to accurately analyze the LTR-RTs insertion burst events in the *B. rapa* genome, which are considered to be the driving force of WGT. To investigate the evolutionary dynamics of LTR-RTs, the insertion time in *B. rapa* "Longyou-7," *B. rapa* "Chiifu," *B. oleracea* (Parkin et al., 2014), and *B. napus* was estimated (**Figure 3B**). The density plot indicated that the *B. rapa* "Longyou-7" genome has a comparatively highest proportion of recent insertions, which is considered to contribute to the larger genome size (413.27Mb) than *B. rapa* "Chiifu" (353.14Mb) (Wang et al., 2011). The LTR-RTs density analysis shows the general existence of LTR-RTs insertion event within all the chromosomes in *B. rapa* with two blocks presents extra high LTR-RTs density, at the end of chromosome A03 and the site about 24 Mb on chromosome 10 (**Figure 3C**). Large scale block with high LTR-RT density was also discovered in chromosome 1 and chromosome 9 at the 16Mb-28Mb and 30Mb-50Mb regions respectively (**Figure 3C**). Another peak in LTR-RTs insertion density occurred about 1 million years ago and can be seen within both the "Chiifu" and "Longyou-7" genomes, which may suggest an ancient LTR insertion event. The activity of transposable elements (TEs) including LTR-RTs is thought to cause various genetic diversities, including the adaption of the environment (Wu et al., 2020).

Graph-Based Pan-Genome

Compared with the reference genome (Chiifu), the genome of LY7 contains 2,174,693 SNPs and 580,073 Indels, involving 4,381,775 bp sequences. But this is far from representing the genetic diversity of species in *B. rapa*. A pan-genome represents an approximation of the entire gene repertoire and provides an important resource for the identification of genetic variants, particularly for larger structural variants such as presence/absence variants (PAVs) and copy number variants (CNVs) of a species. A graph-based pan-genome uses substitutable sequences in a population to represent the variants present at each locus (Li et al., 2022), which can be visualizable and facilitate fast and accurate identification and genotyping of larger SVs within and close to specific genes. To capture the entire genomic diversity, we constructed a high-resolution graph-based pan-genome of *B. rapa* using the assembled genome of "LY7" together with 22 previously reported *B. rapa* genomes representing different morphotypes and ecotypes (Cai et al., 2021) (**Supplementary Table S3**). The *B. rapa* genome of "Chiifu" was used as the reference genome and the other 22 genomes including LY7 were iteratively aligned to the genome of "Chiifu" using Minigraph (Li et al., 2020) (**Supplementary Table S3**). A graphical pan-genome that contained a total of 91,308 structural variations was finally obtained. The size of the pan-genome gradually increased as the number of genomes increased, and eventually leveled off at about 800Mb, indicating that the pan-genome is close to saturation (**Figure 4A**). The graphical pan-genome contains 642,697 sequence fragments, of which 172,811 are from the reference genome, and the rests are PAV identified from other 22 *B. rapa* genomes.

SVs Related to Cold Tolerance

Cold regulated (COR) genes have been isolated and identified from *Arabidopsis thaliana*, canola, rice, and other plants (Hajela et al., 1990; Thomashow et al., 1997; Park et al., 2015). The COR genes encode various functional proteins to resist cold stress and improve cold resistance, including *DREB*, *CBF*, *NAC*, *MYB*,

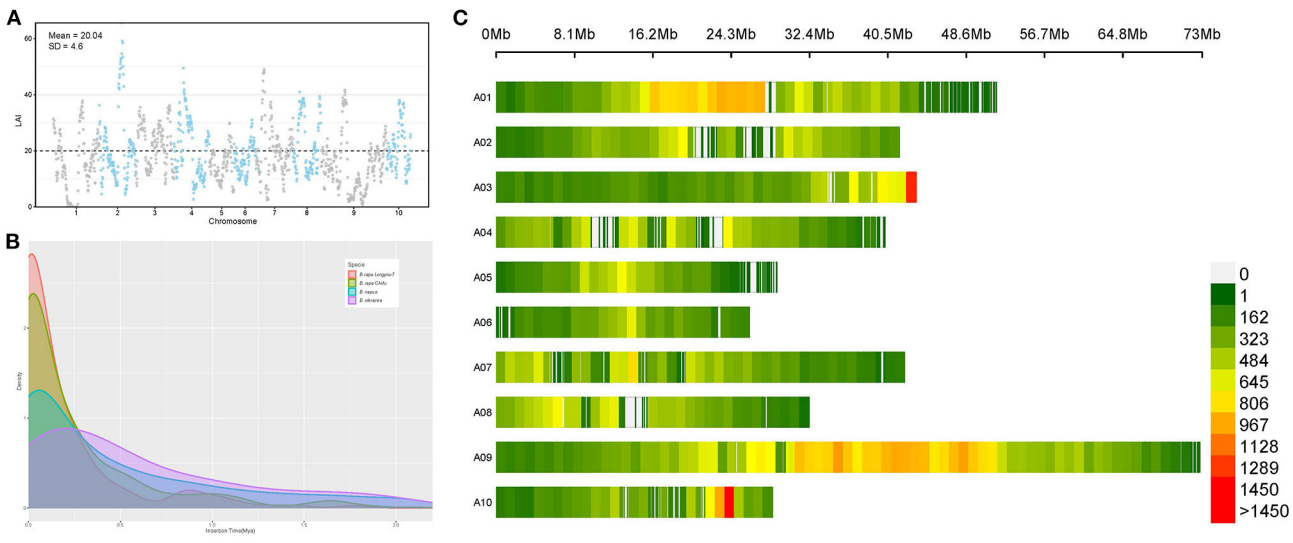


FIGURE 3 | LTR insertion events in *B. rapa* "Longyou-7" genome. **(A)** LAI score reflects the quality of repetitive sequences of *B. rapa* "Longyou-7" genome assembled, with mean LAI = 20.04, SD = 4.6; **(B)** Insertion time distribution of LTR retrotransposons in *B. napus*, *B. oleracea*, *B. rapa*, and *B. rapa* "Longyou-7" genomes; **(C)** The distribution of LTR-RTs density in each chromosome of *B. rapa* "Longyou-7".

bZIP, and *WRKY* (Jaglo-Ottosen et al., 1998; Yoo et al., 2007; Liao et al., 2008, p. 62; Kim et al., 2016; Liu et al., 2019). To reveal the genetic basis of the strong cold tolerance of LY7, we collected a total of 97 COR genes from literature, including 29 genes from Arabidopsis, 13 genes from rice, three genes from *B. rapa,* and 53 genes from *B. napus* (**Supplementary Table S1**). BLAST analysis using amino acid sequences of these genes as query identified 53 orthologous genes in the LY7 genome (**Supplementary Table S1**). Of the 53 COR genes, 40 were found to be multiple copy genes (**Supplementary Tables S4, S5**), which may be related to the WGT event mentioned earlier. Compared to the reference genome of Chiifu, we identified copy number variations (CNV) in 39 out of the 40 multiple copy genes. In addition, with the assistance of the graph-based pan-genome, we found SVs within or upstream (<2 kb) of 17 out of the 40 COR genes. For example, *TCP4* in Arabidopsis encodes a transcription factor that coordinates growth processes during leaf development (Martín-Trillo and Cubas, 2010), which is mediated by miR319. miR319 and its target *TCP4* can act as switches that turn on secondary cell wall synthesis, which is reported to be associated with cold tolerance in plants (Zeng et al., 2018). *Bra032365* is an orthologous gene of *TCP4* in *B. rapa*. It has only one copy in Chiifu, but three copies in LY7 (**Supplementary Table S5**). BLAST search in the graph-based pan-genome identified CNVs and SVs of *Bra032365* in all the 23 *B. rapa* genomes (**Figure 4B**). *Early Responsive to Dehydration 10* (*ERD10*) is reported to play role in the protection of the plants from various stresses, including cold and dehydration (Kim and Nam, 2010). Two orthologous copies of *ERD10*, *Bra012230,* and *Bra025819* located on different chromosomes, were identified in Chiifu, while three copies were identified in LY7. In addition, SVs were identified in both *Bra012230* and *Bra025819* (**Figure 4C**). The *ethylene response factor* (*ERF*) gene family encodes plant-specific transcription factors. *Bra035732* is an orthologous gene of *ERF105*, which was reported to play role in freezing tolerance and cold acclimation of *A. thaliana* (Bolt et al., 2017). Large SVs were detected upstream of *Bra035732* (**Figure 4D**). The CNVs and SVs identified in these COR genes may be associated with cold tolerance in *B. rapa*.

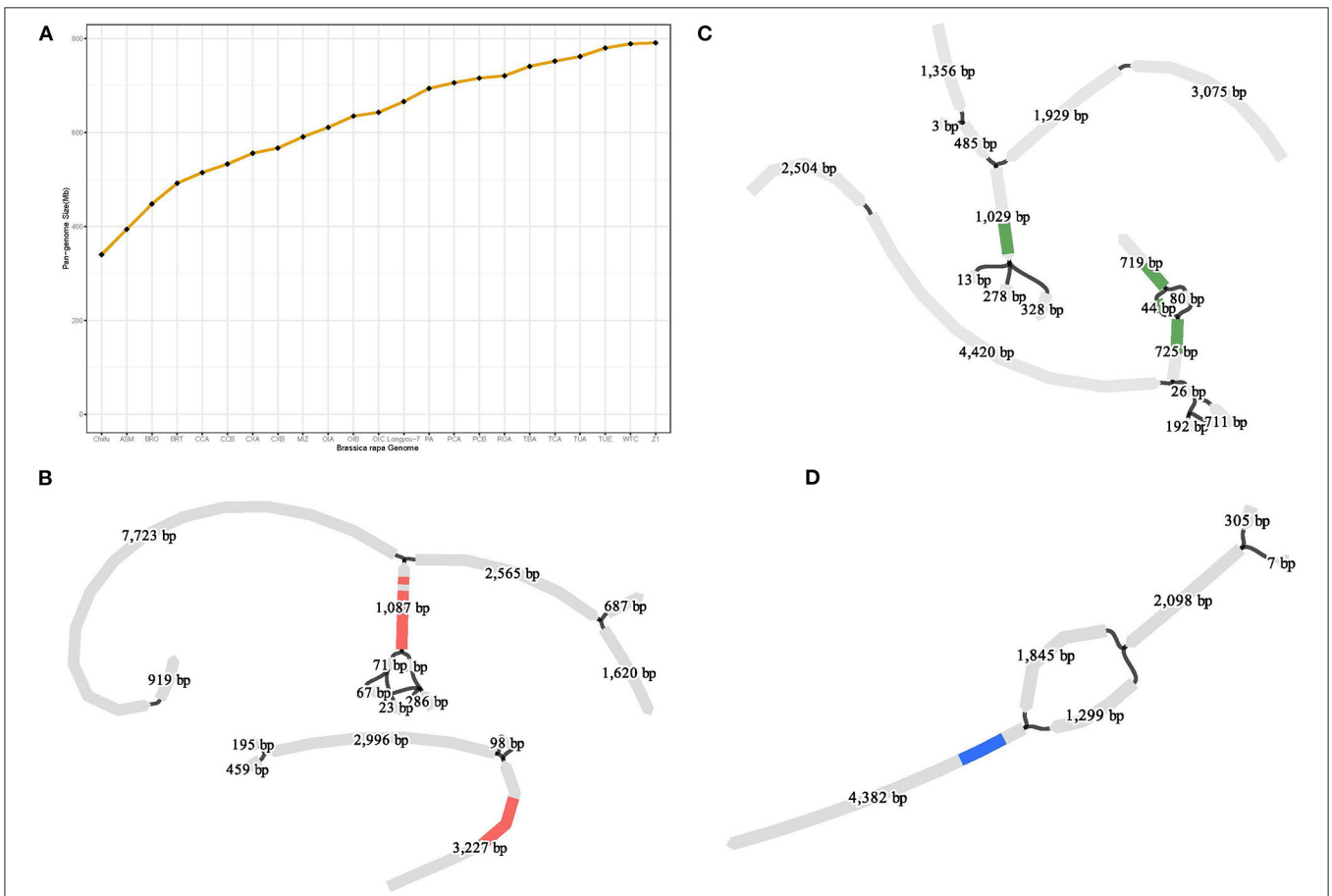


FIGURE 4 | Graph-based pan-genome constructed with 23 *B. rapa* genomes currently exist. **(A)** Pan-genome size increased with more genomes were used; **(B)** Two copies of gene *TCP4* were detected in pan-genome and the SVs around them; **(C)** Two copies of the gene *ERD10* detected in the pan-genome we constructed and the SVs located inside one of the copies **(D)** Structural variants detected in the pan-genome upstream of the gene *ERF105*.

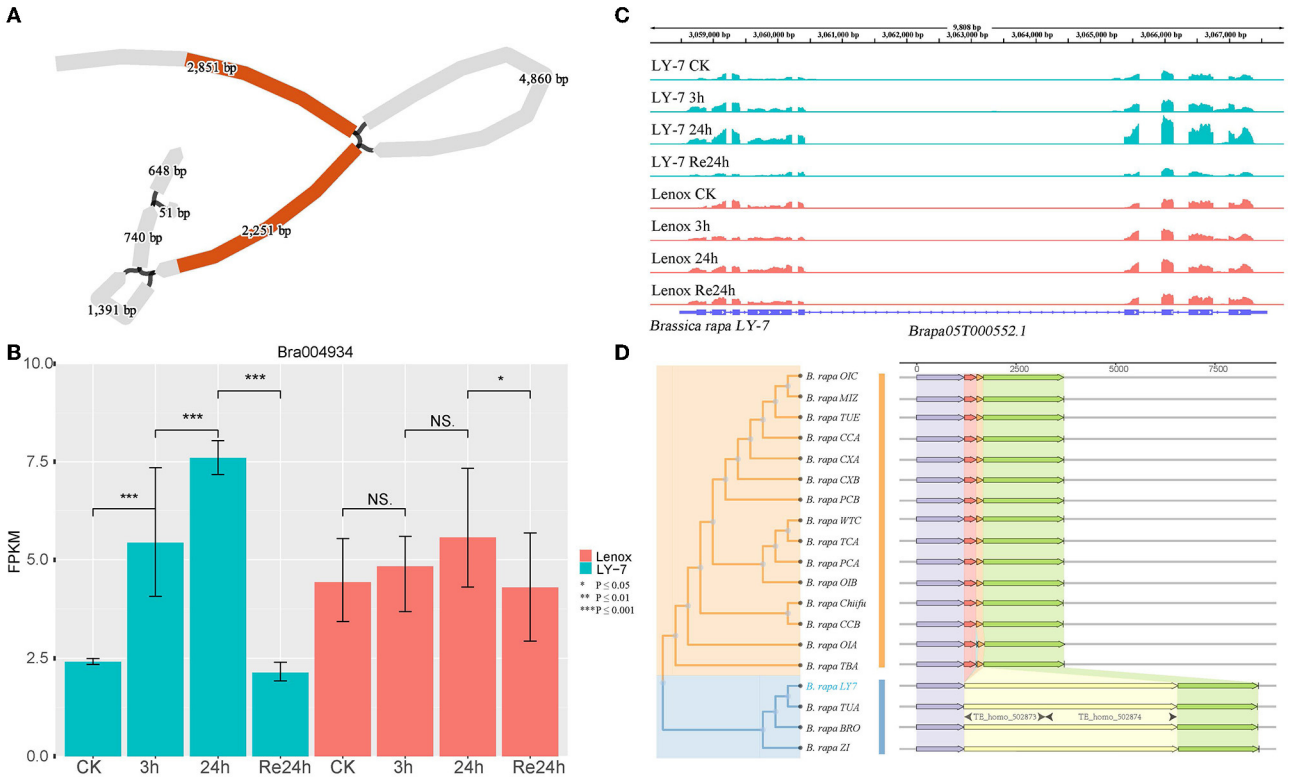


FIGURE 5 | Structure variant of *HDG1* in *B. rapa* "Longyou-7" genome. **(A)** BLAST hit of gene *HDG1* in the pan-genome and 4860bp of structural variation detected within it; **(B)** The differential expression of *HDG1* between "Longyou-7" and "Lenox" within 48 h; **(C)** Coverage of RNAseq reads in the region of gene *HDG1* in "Longyou-7" genome and their abundance; **(D)** Neighbor-joining tree clustering and the gene structure characteristics of gene *HDG1*.

The Potential Impact of SVs Related to Cold Tolerance

To further reveal if the CNVs and SVs identified in the 40 COR genes are associated with cold tolerance, we analyzed the differences in expression levels of these genes before and after cold treatment. Plants of LY7 and Lenox grow in a 22 °C growth chamber (with 16/8h light/dark cycle) to a six-leaf stage and then were transferred to a 4 °C growth chamber for 3 and 24 h, and then recovered at 22 °C for 24h. Growth points of these plants were collected for RNA-seq (Ma et al., 2019). Of the 17 genes having structural variations within coding sequence regions (CDS) or upstream regulatory sequences, two genes, *HDG1* and *BrANS3*, displayed different expression patterns between LY7 and Lenox, while the other genes showed similar expression patterns in LY7 and Lenox before and after cold treatment. *HDG1*, encoding a protein in the homeodomain (HD)-START transcription factor family also known as the Class IV Homeodomain-Leucine Zipper transcription factor family (Nakamura et al., 2006), was previously reported to assist with the growth and development of plants under unfavorable environments including cold stress (Sharif et al., 2021). Overexpressing *HDG1* in plants significantly influenced the root system and improved the resistance to stresses (Horstman et al., 2015). *Bra004934* is a homologous gene of *HDG1*. With the graph-based pan-genome, we identified a 4860 bp LTR inserted in the first intron of *HDG1* in "LY7" (**Figure 5A**). Alignment of RNA-seq reads to the LY7 genome indicated that the insertion of LTR in the first intron did not change the protein-coding sequence and disrupted the transcription of *HDG1* (**Figure 5C**). Further comparative analysis identified the same LTR in *HDG1* of BRO (Broccolieto), TUA (Turnip), and Z1 (sarson type) (Cai et al., 2021) genomes (**Figure 5D**). The expression level of *HDG1* in "Longyou-7" increased 2 folds after 3h cold treatment and four folds after 24h cold treatment ($P = 0.01$). Its expression level was restored to the untreated level after recovery at 22°C for 24h (**Figures 5B,C**). Whereas the expression level of *HDG1* in Lenox did not show a significant difference before and after cold treatment (**Figures 5B,C**). Thus, the differential expression of *HDG1* in "LY7" is thought to be associated with the cold-tolerance, and the LTR inserted in the first intron of *HDG1* might be cold-inducible. The structural variation in *HDG1* was further examined in 80 samples from the resequencing dataset of spring Chinese cabbage populations reported in previous studies (Su et al., 2018), and the results showed that the structural variation was detected in 82.5% of the samples, indicating that the structural

variation is widespread in cold-tolerant spring cabbage varieties (**Supplementary Table S6**).

Anthocyanins are water-soluble flavonoid pigments widely distributed in the petals, fruits, stems, and leaves of plants (Potapovich and Kostyuk, 2003; Lo Piero et al., 2005; Yuan et al., 2009). In addition, anthocyanins are natural antioxidants that can strongly scavenge free radicals and reactive oxygen species (ROS) (Gould, 2004). Anthocyanidin synthase (ANS) is one of the structural genes encoding multiple enzymes in anthocyanin biosynthesis. *BrANS3* (*Bra013652*) was reported to be strongly associated with anthocyanin accumulation and resistance to cold stress of *B. rapa* (Ahmed et al., 2015). In the constructed graph-based pangenome, we detected structural variations within this gene (**Figure 6A**). A *Mariner* insertion of about 240 bp was identified in 10 genomes including OIA, CCA, Z1, OIC, OIB, WTC, MIZ, CXB, TBA, and PCA, but not in the other genomes including the two-known cold-tolerant varieties "LY7" and "Chiifu" (**Figure 6B**). Due to the SNPs in this structural variation, four branches representing four different allelic genotypes can be observed in the graph-based pan-genome (**Figure 6A**). *BrANS3* was differentially expressed between "LY7" and "Lenox" varieties before cold treatment. The expression level of *BrANS3* was very low in Lenox before cold treatment and not elevated after cold treatment, while the expression level in LY7 was significantly increased after 24h cold treatment (**Figure 6C**). We further demonstrated that this gene was differentially expressed in cold-tolerant and cold-sensitive varieties by using the genome of the "TBA" variety, which has the *Mariner* insertion of ∼240 bp in *BrANS3*, as the reference genome for transcriptome mapping, and found that the insertion did not change the protein-coding sequence of this gene (**Figures 6C,D**). Therefore, we speculate that the insertion of this fragment may be widely present inside the *BrANS3* of cold-sensitive rapeseed varieties, hindering the expression of the gene and thus causing rapeseed to exhibit cold-sensitive traits, while *BrANS3* without the insertion can express normally and improve the cold tolerance of rapeseed. Alignment of the 80 resequencing data mentioned earlier (Su et al., 2018) to the TBA variety genome revealed that the structural variant was not widespread in the spring Chinese cabbage population and was detected in only 12.5% of the samples (**Supplementary Table S7**). A weighted gene co-expression network analysis (WGCNA) was performed on the transcriptome data (Ma et al., 2019), and the top 100 co-expressed genes with weight values were screened, five of which contained cold response terms in the GO annotation information (Ma et al., 2019). Their interactions were predicted in the STRING database, and from the results, three of them are located upstream of gene *BrANS3* and may play a regulatory role in the expression of *BrANS3*.

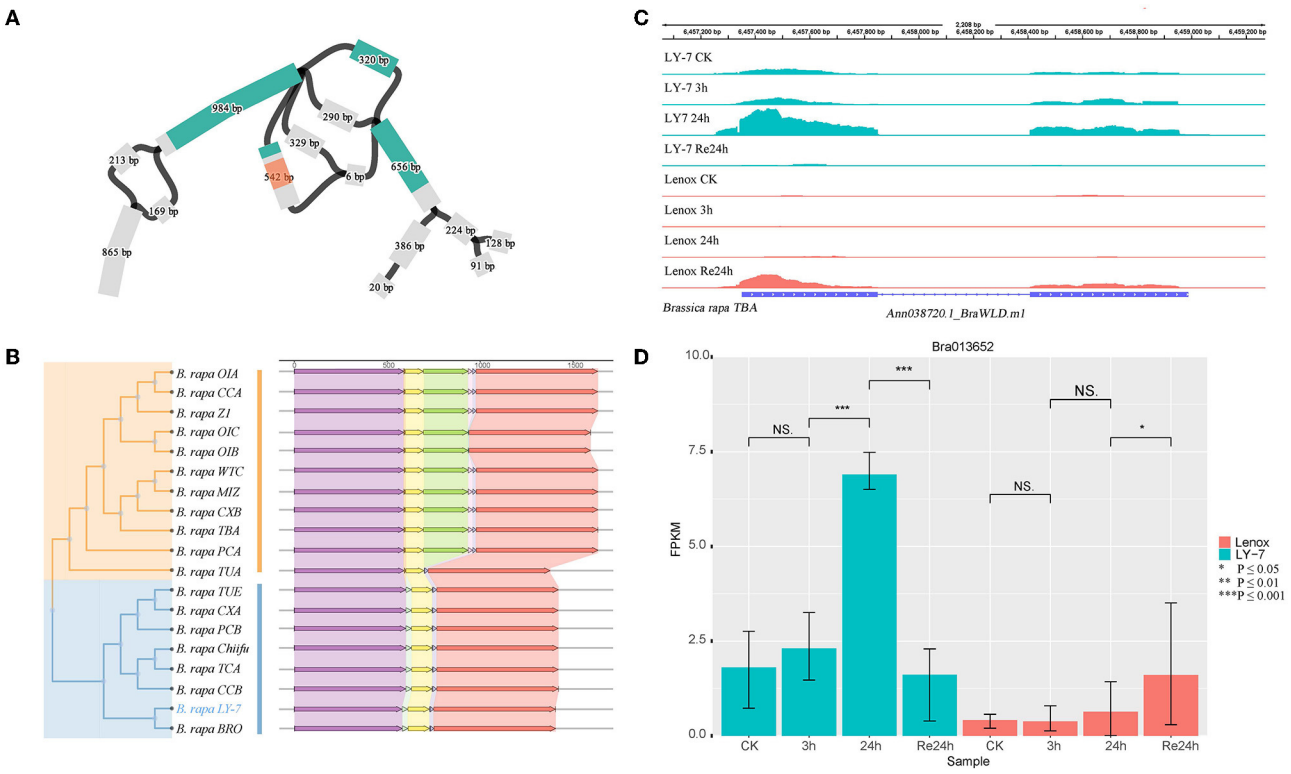


FIGURE 6 | Structure variant of *BrANS3* in *B. rapa* "Longyou-7" genome. **(A)** BLAST hit of *BrANS3* gene of "Longyou-7" genome (blue) and a 240 bp of structural variation (red transparent) detected within the graph-based pan-genome; **(B)** Neighbor-joining tree clustering and the gene structure characteristics of *BrANS3* gene; **(C)** Coverage of RNA-Seq reads in the region of *BrANS3* gene in "TBA" (From Tibet, China) genome and their abundance; **(D)** The differential expression of *BrANS3* between "Longyou-7" and "Lenox" cultivars within 48 h.

DISCUSSION

In this study, we sequenced and assembled for the first time a high-quality chromosome-level genome of rapeseed cold-tolerant variety "Longyou-7" and performed a series of correlation analyses based on the assembled genome to elucidate cold-tolerance traits in this cultivar. Comparative genomic analysis of the "Longyou-7" genome with some other cruciferous plant genomes identified the whole-genome triplication event of *B. rapa* in the "Longyou-7" genome. Analysis of TEs revealed a recent LTR insertion event in the "Longyou-7" genome, which is thought to be possibly related to cold adaption in this variety of rapeseed. Together with the currently existing 22 *B. rapa* genomes, we build a graph-based pan-genome with high resolution. Subsequently, we collected 97 genes (**Supplementary Table S1**) that have been reported to be associated with cold stress located in rice, kale, oilseed rape, and Arabidopsis, and extracted their homologs in rapeseed based on the synteny relationship between these genomes of these species and rapeseed genomes, and further looked for SVs within these genes with the constructed pan-genomes. The transcriptome data of "Longyou-7" reported in previous studies with Lenox (Ma et al., 2019) were used to see the expression of homologous genes with structural variants. We detected SVs in 16 cold tolerance-related homologs (**Supplementary Table S4**) in the "Longyou-7" genome and two of them, *HDG1* and *BrANS3,* were differentially expressed before and after cold treatment between "Longou-7" and "Lenox" varieties. Among them, *HDG1* was previously reported to enhance plant tolerance to cold stress (Horstman et al., 2015), and in this study, we detected a 4,860 bp structural variation in the internal intron region of this gene in the cold-tolerant "Longyou-7" variety, which was further confirmed to be due to the insertion of two Copia LTR retrotransposons. *BrANS3* was previously reported to be strongly associated with cold tolerance in *B. rapa* (Ahmed et al., 2015), and here a deletion of about 240 bp was detected in the internal intron region of this gene within both known cold-tolerant species, "Longyou-7" and "Chiifu." We further supported these two findings by sequence-structure analysis. Transcript coverage showed that the structural variation in both genes did not affect their protein-coding sequences, and differential expression of both genes was detected between cold-tolerant and cold-sensitive varieties. Therefore, the structural variants detected in the cold-tolerance-related genes *HDG1* and *BrANS3* are thought to have led to their differential expression between cold-tolerant and cold-sensitive varieties, and consequently to the cold-tolerant phenotype exhibited by the "Longyou-7" variety.

Collectively, our research provides a high-quality genomic resource for the study of the cold-tolerant phenotype in *B. rapa*. Results of the comparative genomic analysis confirmed the WGT event experienced by *B. rapa* "Longyou-7" genome. Moreover, the constructed graph-based pan-genome could serve as a resource for the excavation of cold tolerance-related SVs of *B. rapa*. Furthermore, we uncovered two possible SVs that contribute to the cold-tolerant trait of *B. rapa* variant "Longyou-7," which reveals new insights into the cold tolerance of rapeseed.

MATERIALS AND METHODS

Plant Materials

"Longyou-7" is a winter turnip rape with super freeze-tolerance bred by scientists at Gansu Agricultural University (Lanzhou, China). They crossed a landrace with strong cold resistance, Chenjiazui winter turnip rape, with high-yielding Tianyou 4 with weak cold resistance in 1996 (Sun et al., 2011). The F2, F3, and higher generations family lines were grown at multiple locations around Lanzhou city (latitude: 34°33′-39°46′N; longitude: 103°.82′E; Altitude:1,083–1,477 m,) under natural conditions in winter. Families with overwinter survival rate higher than 80% were selected for each generation. One of the family lines with super cold resistance was registered as "Longyou-7" in 2007, which can survive in the winter in the region between N34° and N48° and under as low as−31.9 °C environment with a survival rate higher than 80% (Zhou et al., 2014). Lenox is a freeze-sensitive variety also bred by scientists at Gansu Agricultural University. It has a overwinter survival rate of 0-10% at Tianshui (34°36′ N, 105°39′ E, the annual average temperature is 12.9°C), Huining (35°89′ N, 104°62′ E, the annual average temperature is 9.2°C) and Qingyang (35°38′ N, 107°35′ E, the annual average temperature is 10°C).

We also measured the activity of other ROS scavenging enzymes, including superoxide dismutase (SOD), catalase (CAT), glutathione reductase (GR), and peroxidase, in the ospp18 mutant and wild-type plants. Among these ROS scavenging enzymes, the activity of GR was significantly lower in the ospp18 mutant than in the wild type under both the drought stress and non-stress conditions (**Figure 6B**).

Measurements of Phenotypic and Physiological Traits

Longyou 7 and Lenox were sowed at the experimental farm of Gansu Agricultural University, Yongdeng County, Gansu, on August 20, 2020. Field management was conducted as regular agricultural practice. The rapeseed leaves and roots were collected when growth reached the five-leaf stage (October).

Leaves were collected every 1 month after sowing from plants grown in the field until December. Root shoot ratio refers to the ratio of fresh weight or dry weight of the underground part and aboveground part of rapeseed. Oxidative damage was estimated by measuring the content of MDA as described previously (Du et al., 2010). The content of soluble protein (SP) was measured as previously described (Arminian and Dehghani Bidgoli, 2019). The activity of ROS scavenging enzymes, such as ascorbate peroxidase (APX) activity, peroxidase (POD) activity (Quiroga et al., 2000), superoxide dismutase (SOD) activity, catalase (CAT) activity, were measured as described in previous studies (Quiroga et al., 2000; Yao et al., 2011; Zhang et al., 2013).

Library Construction and Sequencing

Genomic DNA was extracted following the method by which genomic DNA was extracted for Illumina and Pacbio library construction and sequencing. Libraries for Pacbio-HiFi sequencing were constructed using the SMRTbell Express Template Prep Kit v2.0 following the manufacturer"s protocols

provided by PacBio Company and sequenced on the PacBio platform. The created Hi-C library was digested into units with Dpn II and sequenced by Illumina HiSeq 4000 platform with 150 bp reads length with a 200 bp insert size and sequenced on Illumina platform. Library for Illumina pair-end genome sequencing was built following the provided standard protocol.

Genome Assembly

The *B. rapa* "Longyou-7" genome was *de novo* assembled using the PacBio SMRT data. Falcon was used for performing subreads polishing and contigs assembly (falcon−201711.02−16.04 - py 2.7), and parameter length_cut_off_pr was set at 6,000. Canu v1.6 (Koren et al., 2017) was used for the assembly of subreads polished by Falcon (Chin et al., 2016), and parameter correctedErrorRate was set at 0.05. Then the PacBio reads were mapped to the draft contigs by pbalign, whose result was polished with Quiver (Chin et al., 2013) using the arrow algorithm. The gained contigs were polished with Illumina PE reads (insertion size 350bp) by pilon1.18. Unique sequences generated by Canu were not found in Falcon assembly. Hi-C reads were used for scaffolding the draft assembly genome using the 3D-DNA pipeline (Dudchenko et al., 2017), with the parameter -i set as 1 and -r set as 5. Hi-C reads were aligned to the polished contigs following the reported Juicer pipeline (Durand et al., 2016b). The result of 3D-DNA was polished using Juicebox (Durand et al., 2016a). We finally got 10 chromosome-length scaffolds. The completeness and accuracy of *B. rapa* 'Longyou-7' assembly were assessed using BUSCO (Seppey et al., 2019) with the embryophyte_odb10 dataset.

Repeat Element Annotation

An extensive *de-novo* TE Annotator (EDTA) pipeline (Ou et al., 2019) was used for building a whole-genome *de novo* repeat library and performing the identification of the transposable elements (TEs). The *de novo* detection of long terminal repeat (LTR) retrotransposons in *B. rapa* "Longyou-7" genome we detected was performed with LTR_Finder (Xu and Wang, 2007) and LTRharvest (Ellinghaus et al., 2008), the results of which were further filtered with LTR_retriever (Ou and Jiang, 2018), and the LTR insertion time was calculated at the same time. The extracted LTRs were used for the calculating LTR Assembly Index (LAI) to evaluate the assembly continuity of *B. rapa* "Longyou-7" genome using LTR-RTs by LAI program (Ou et al., 2018).

Gene Prediction and Annotation

Multiple strategies were used for the prediction of gene structure including homologous prediction, *de novo* prediction, and evidence-based prediction. *De novo* prediction software including Augustus (Stanke et al., 2006), GlimmerHMM (Majoros et al., 2004), and SNAP (Söllner et al., 1993). Genewise (Birney et al., 2004) was used for the homologous prediction of gene structure. RNA-Seq data of mixed tissue were used for evidence-based prediction by EVidenceModeler (Haas et al., 2008). The correction and addition of UTR, variable clipping, and other information were performed with PASA (Haas et al., 2008).

Whole Genome Triplication in *B. rapa* "Longyou-7" Genome

To study the evolution of *B. rapa* "Longyou-7" genome, syntenic blocks between *B. rapa* "Longyou-7" and *A. thaliana* were defined and presented by the Mcscan python version (Tang et al., 2008) with default parameters. The analysis revealed long stretches of triplication within the *B. rapa* "Longyou-7" genome assembled, and they were found to be not only inter-chromosomal but also intra-chromosomal. Paralog analysis within *B. rapa* "Longyou-7" genome was performed with reciprocal best hits (RBH) from self-BLASTp using all the primary protein sequences. Self-BLASTp was performed using python script blast_rph.py (https://github.com/peterjc/galaxy_blast/blob/master/tools/blast_rbh/blast_rbh.py) in the galaxy_blast package (Cock et al., 2015). Further analysis of the synonymous substitution rate (Ks) of RBH gene pairs was calculated based on the MA model by KaKs_Calculator v2.0 (Wang et al., 2010) and ParaAT2.0 (Zhang et al., 2012) to confirm the WGT event *B. rapa* "Longyou-7" genome underwent and the time it happened and 22,078 RHB paralogous gene pairs in the *B. rapa* genome. Both the WGT peak in *B. rapa* "Longyou-7" genome and *B. rapa* current reference genome and the peak indicates the divergence between *B. rapa* and other *Brassica ssp.* was detected. Single copy ortholog protein sequences among *A. thaliana* (Sloan et al., 2018), *Capsella rubella* (Slotte et al., 2013), *Thlaspi arvense* (Dorn et al., 2015), *B. oleracea* (Parkin et al., 2014), *B. napus* "ZS11" (Song et al., 2020), *B. rapa* "Chiifu" (Wang et al., 2011) and *B. rapa* "Longyou-7" was generated with OrthoFinder (Emms and Kelly, 2019), and then aligned with MUSCLE (Edgar, 2004) based on the alignment result phylogenetic tree was constructed using RAxML (Stamatakis, 2014) and r8s (Sanderson, 2003) was used for estimating of the split times. The species phylogenic tree of the 18 reported *B. rapa* genomes (Cai et al., 2021) and our "Longyou-7" assembly was constructed using the same process, with species *B. oleracea* being used as the outgroup. R package "ggtree" was used for phylogenetic tree mapping and landscaping.

RNA-seq Data Analysis

RNAseq data set was obtained from the published research (Ma et al., 2019) (NCBI accession number: SRP179662). A quality check was performed with FastQC. Trimmomatic (Bolger et al., 2014) was used for filtering reads with low quality to gain clean reads. Clean reads were aligned to *B. rapa* reference genome (Wang et al., 2011) using STAR (Dobin et al., 2013), and samtools was used to transform SAM file into BAM file. The count of reads mapped to the reference genome was counted using RSEM (Li and Dewey, 2011). Cufflinks (Trapnell et al., 2010) was used to calculate fragments Per Kilobase of exon model per Million mapped fragments (FPKM). Differential expression analysis was performed with "DESeq2." The statistical analysis and mapping of expression levels of gene *HDG1* and *BrANS3* were performed using the R packages "ggplot2" and "ggsignif," and the comparison of reads on the reference genome was further viewed and analyzed using igvtools (v2.11.2).

Graph-Based Pan-Genome Construction

Twenty-two previously reported *B. rapa* genomes were downloaded from the Brassicaceae Database (BRAD) (http://www.brassicadb.cn/) and National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) (Belser et al., 2018; Zhang et al., 2018; Cai et al., 2021; Li et al., 2021), including heading Chinese Cabbage, turnips (Chinese and European turnips), sarsons (sarson, rapid cycling, and oilseed), pak choi (pak choi, wutacai, and caixin), and Japanese morphotype (mizuna). Of these genomes, the annotation information of four genomes (ASM, BRT, PA, ROA) is not available. The 22 previously reported *B. rapa* genomes and the *B. rapa* "Longyou-7" genome were built into a variant graph using minigraph following the published process with default parameters (Li et al., 2020). The genome of variety "Chiifu" (Wang et al., 2011) was used as the reference genome, and the genomes of the other 22 varieties were mapped iteratively, one by one, using in-house shell scripts, for sequence-to-graph mapping, and the generated GFA format file was used to call bubbles, which represent structural variants, using gfatools (v0.4-r214-dirty) (https://github.com/lh3/gfatools). The constructed graph-based pan-genome in GFA format was visualized using Bandage (v0.8.1) (Wick et al., 2015).

Identification of SVs Related to Cold Tolerance

To find the SVs related to the cold tolerance phenotype of *B. rapa* "Longyou-7", 97 cold tolerance related genes from *B. rapa*, *B. oleracea*, *A. thaliana,* and rice (*Oryza sativa*) (**Supplementary Table S1**) was collected and aligned to19 genomes including 18 *B. rapa* genomes reported in the previous study (Cai et al., 2021) and "Longyou-7" genome with Mcscan python version (Tang et al., 2008). Homologous genes identified in these *B. rapa* genomes were aligned to the graph-based pan-genome to check if there are SVs within or around them. Here we selected *HDG1* (gene ID: *Bra004934*) and *BrANS3* (gene ID: *Bra013652*) to perform further checks. The longest transcript of it was extracted according to the GFF file with an in-house bash script. The inter-genome genes relationship was built with the MCScan python version (Tang et al., 2008). Sequences of *HDG1* homologous genes in the 18 *B. rapa* genome were extracted with in-house scripts. The synteny blocks were analyzed using Mauve (v2015-02-26) (Darling et al., 2004) and presented with the R package "gggenes." The neighbor-joining tree of sequences of each gene was constructed with MEGA11 (Tamura et al., 2021) with default parameters and drawn using the R package "ggtree." Information on transposable elements located in the *HDG1* of the "Longyou-7" genome was obtained from the output of the EDTA pipeline (Ou et al., 2019).

Gene Coexpression Network

Gene co-expression network analysis was performed using transcriptomic data of 41,019 genes with the R package "WGCNA" (Langfelder and Horvath, 2008). Terrain Overlap Matrix (TOM) was calculated with the "blockwiseModules" function in the WGCNA package, where the TOMType was set to "unsigned," and the minimum module size was set to 30. The co-expression network was constructed by a one-step method, maxBlockSize was set to 50000. From the results, all genes interacting with the *BrANS3* were extracted, ranked by weight value, and the top 100 were selected for further analysis, and the interactions between the proteins encoded by these genes were predicted using the STRING database (https://string-db.org) (Szklarczyk et al., 2021). The visualization of the interaction network was implemented using Cytoscape (v3.9.0) (Shannon et al., 2003).

Alignment of Resequencing Data

Eighty population resequencing data were obtained from the dataset of previous studies and matched to the corresponding reference genomes using BWA (v0.7.17-r1188) (Li, 2013). The alignment results were converted to BAM format files using samtools (v1.9) (Li et al., 2009) and sorted. Finally, the sorted BAM files were counted using bedtools (v2.30.0) (Quinlan and Hall, 2010).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI, SRR18959686.

AUTHOR CONTRIBUTIONS

JW and WS conceived and designed the study. X-DX analyzed the data. LL, LM, YP, WW, X-YH, J-MS, GL, YF, and XL contributed materials and analysis tools. JW and X-DX wrote the paper. KL revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by the Research Program Sponsored by State Key Laboratory of Aridland Crop Science, Gansu Agricultural University (No. GSCS-2020-Z1) and the China Agriculture Research System of MOF and MARA (CARS-12).

ACKNOWLEDGMENTS

The authors thank Yu-yu Zheng for the helpful discussion during data analysis and visualization. The authors thank all the individuals who participated in the study, as well as all the funds that supported this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.936958/full#supplementary-material

REFERENCES

Ahmed, N. U., Park, J. I., Jung, H. J., Hur, Y., and Nou, I. S. (2015). Anthocyanin biosynthesis for cold and freezing stress tolerance and desirable color in Brassica rapa. *Funct. Integr. Genomics* 15, 383–394. doi: 10.1007/s10142-014-0427-7

Apel, K., and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701

Arminian, A., and Dehghani Bidgoli, R. (2019). Simultaneous responses of photosystem II and soluble proteins of rapeseed to cold acclimation. *Cell. Mol. Biol. Noisy Gd. Fr.* 65, 37–49. doi: 10.14715/cmb/2019.65.2.7

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, 879–887. doi: 10.1038/s41477-018-0289-4

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. *Genome Res.* 14, 988–995. doi: 10.1101/gr.1865504

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolt, S., Zuther, E., Zintl, S., Hincha, D. K., and Schmülling, T. (2017). ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. *Plant Cell Environ.* 40, 108–120. doi: 10.1111/pce.12838

Cai, C., Wang, X., Liu, B., Wu, J., Liang, J., Cui, Y., et al. (2017). Brassica rapa Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. *Mol. Plant* 10, 649–651. doi: 10.1016/j.molp.2016.11.008

Cai, X., Chang, L., Zhang, T., Chen, H., Zhang, L., Lin, R., et al. (2021). Impacts of allopolyploidization and structural variation on intraspecific diversification in Brassica rapa. *Genome Biol.* 22, 1–24. doi: 10.1186/s13059-021-02383-2

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

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi: 10.1038/nmeth.4035

Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E., and Soranzo, N. (2015). NCBI BLAST+ integrated into Galaxy. *GigaScience* 4, 39. doi: 10.1186/s13742-015-0080-7

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dorn, K. M., Fankhauser, J. D., Wyse, D. L., and Marks, M. D. (2015). A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 22, 121–131. doi: 10.1093/dnares/dsu045

Du, H., Wang, N., Cui, F., Li, X., Xiao, J., and Xiong, L. (2010). Characterization of the beta-carotene hydroxylase gene DSM2 conferring drought and oxidative stress resistance by increasing xanthophylls and abscisic acid synthesis in rice. *Plant Physiol.* 154, 1304–1318. doi: 10.1104/pp.110.163741

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. doi: 10.1126/science.aal3327

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3, 99–101. doi: 10.1016/j.cels.2015.07.012

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al. (2016b). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9, 18. doi: 10.1186/1471-2105-9-18

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y

Esterbauer, H., Schaur, R. J., and Zollner, H. (1991). Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic. Biol. Med.* 11, 81–128. doi: 10.1016/0891-5849(91)90192-6

Gómez-Campo, C., and Prakash, S. (1999). "2 Origin and domestication," in *Developments in Plant Genetics and Breeding*, eds Gómez-Campo, C. (Elsevier), 4, 33–58. doi: 10.1016/S0168-7972(99)80003-6

Gould, K. S. (2004). Nature's swiss army knife: the diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol.* 2004, 314–320. doi: 10.1155/S1110724304406147

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Hajela, R. K., Horvath, D. P., Gilmour, S. J., and Thomashow, M. F. (1990). Molecular Cloning and Expression of cor (Cold-Regulated) Genes in Arabidopsis thaliana. *Plant Physiol.* 93, 1246–1252. doi: 10.1104/pp.93.3.1246

Horstman, A., Fukuoka, H., Muino, J. M., Nitsch, L., Guo, C., Passarinho, P., et al. (2015). AIL and HDG proteins act antagonistically to control cell proliferation. *Dev. Camb. Engl.* 142, 454–464. doi: 10.1242/dev.117168

Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., and Thomashow, M. F. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* 280, 104–106. doi: 10.1126/science.280.5360.104

Jung, G. A., Shih, S. C., and Shelton, D. C. (1967). Seasonal changes in soluble protein, nucleic acids, and tissue pH related to cold hardiness of alfalfa. *Cryobiology* 4, 11–16. doi: 10.1016/S0011-2240(67)80181-0

Kim, C. Y., Vo, K. T. X., Nguyen, C. D., Jeong, D. H., Lee, S.-K., Kumar, M., et al. (2016). Functional analysis of a cold-responsive rice WRKY gene, OsWRKY71. *Plant Biotechnol. Rep.* 10, 13–23. doi: 10.1007/s11816-015-0383-2

Kim, S. Y., and Nam, K. H. (2010). Physiological roles of ERD10 in abiotic stresses and seed germination of Arabidopsis. *Plant Cell Rep.* 29, 203–209. doi: 10.1007/s00299-009-0813-0

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly *via* adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi: 10.1186/1471-2105-9-559

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi: 10.1186/1471-2105-12-323

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Genomics*. 1303.3997. doi: 10.48550/arXiv.1303.3997

Li, H., Feng, X., and Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. *Genome Biol.* 21, 265. doi: 10.1186/s13059-020-02168-z

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

Li, H., Wang, S., Chai, S., Yang, Z., Zhang, Q., Xin, H., et al. (2022). Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* 13, 682. doi: 10.1038/s41467-022-28362-0

Li, P., Su, T., Zhao, X., Wang, W., Zhang, D., Yu, Y., et al. (2021). Assembly of the non-heading pak choi genome and comparison with the genomes of heading Chinese cabbage and the oilseed yellow sarson. *Plant Biotechnol. J.* 19, 966–976. doi: 10.1111/pbi.13522

Liao, Y., Zou, H. F., Wei, W., Hao, Y. J., Tian, A. G., Huang, J., et al. (2008). Soybean GmbZIP44, GmbZIP62 and GmbZIP78 genes function as negative regulator of

ABA signaling and confer salt and freezing tolerance in transgenic Arabidopsis. *Planta* 228, 225–240. doi: 10.1007/s00425-008-0731-3

Liu, C., Schläppi, M. R., Mao, B., Wang, W., Wang, A., and Chu, C. (2019). The bZIP73 transcription factor controls rice cold tolerance at the reproductive stage. *Plant Biotechnol. J.* 17, 1834–1849. doi: 10.1111/pbi.13104

Lo Piero, A. R., Puglisi, I., Rapisarda, P., and Petrone, G. (2005). Anthocyanins accumulation and related gene expression in red orange fruit induced by low temperature storage. *J. Agric. Food Chem.* 53, 9083–9088. doi: 10.1021/jf051609s

Ma, L., Coulter, J. A., Liu, L., Zhao, Y., Chang, Y., Pu, Y., et al. (2019). Transcriptome Analysis Reveals Key Cold-Stress-Responsive Genes in Winter Rapeseed (Brassica rapa L.). *Int. J. Mol. Sci.* 20, E1071. doi: 10.3390/ijms20051071

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinforma. Oxf. Engl.* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315

Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K. T., et al. (2006). Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. *Plant Physiol.* 141, 1363–1375. doi: 10.1104/pp.106.077388

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 46, e126. doi: 10.1093/nar/gky730

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20, 275. doi: 10.1186/s13059-019-1905-y

Park, S., Lee, C.-M., Doherty, C. J., Gilmour, S. J., Kim, Y., and Thomashow, M. F. (2015). Regulation of the Arabidopsis CBF regulon by a complex low-temperature regulatory network. *Plant J. Cell Mol. Biol.* 82, 193–207. doi: 10.1111/tpj.12796

Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.* 15, R77. doi: 10.1186/gb-2014-15-6-r77

Potapovich, A. I., and Kostyuk, V. A. (2003). Comparative study of antioxidant properties and cytoprotective activity of flavonoids. *Biochem. Biokhimiia* 68, 514–519. doi: 10.1023/A:1023947424341

Prakash, S., Wu, X., and Bhat, S. R. (2011). History, evolution, and domestication of brassica crops. *Plant Breed. Rev.* 35, 19–84. doi: 10.1002/9781118100509.ch2

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Quiroga, M., Guerrero, C., Botella, M. A., Barceló, A., Amaya, I., Medina, M. I., et al. (2000). A tomato peroxidase involved in the synthesis of lignin and suberin. *Plant Physiol.* 122, 1119–1127. doi: 10.1104/pp.122.4.1119

Ramchiary, N., and Lim, Y. P. (2011). "Genetics of Brassica rapa L.," in *Genetics and Genomics of the Brassicaceae Plant Genetics and Genomics: Crops and Models,* eds. Schmidt, R., and Bancroft, I. (New York, NY: Springer), 215–260.

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinforma. Oxf. Engl.* 19, 301–302. doi: 10.1093/bioinformatics/19.2.301

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol. Clifton NJ* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14

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, 2498–2504. doi: 10.1101/gr.1239303

Sharif, R., Raza, A., Chen, P., Li, Y., El-Ballat, E. M., Rauf, A., et al. (2021). HD-ZIP gene family: potential roles in improving plant growth and regulating stress-responsive mechanisms in plants. *Genes* 12, 1256. doi: 10.3390/genes12081256

Sloan, D. B., Wu, Z., and Sharbrough, J. (2018). Correction of persistent errors in arabidopsis reference mitochondrial genomes. *Plant Cell* 30, 525–527. doi: 10.1105/tpc.18.00024

Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., and Maumus, F., Guo, Y.-L., et al. (2013). The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831–835. doi: 10.1038/ng.2669

Söllner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., et al. (1993). SNAP receptors implicated in vesicle targeting and fusion. *Nature* 362, 318–324. doi: 10.1038/362318a0

Song, J. M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–439. doi: 10.1093/nar/gkl200

Su, T., Wang, W., Li, P., Zhang, B., Li, P., Xin, X., et al. (2018). A Genomic Variation Map Provides Insights into the Genetic Basis of Spring Chinese Cabbage (Brassica rapa ssp. pekinensis) Selection. Mol. *Plant* 11, 1360–1376. doi: 10.1016/j.molp.2018.08.006

Sun, W. C., Wu, J. Y., Fang, Y., Liu, Q., Yang, R. Y., Ma, W. G., et al. (2011). Growth and development characteristics of winter rapeseed northern-extended from the cold and arid regions in China. *Acta Agron. Sin.* 36, 2124–2134. doi: 10.3724/SP.J.1006.2010.02124

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkab835

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Thomashow, M. F., Stockinger, E. J., Jaglo-Ottosen, K. R., Gilmour, S. J., and Zarka, D. G. (1997). Function and regulation of Arabidopsis thaliana COR (cold-regulated) genes. *Acta Physiol. Plant.* 19, 497–504. doi: 10.1007/s11738-997-0046-1

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

Trick, M., Bancroft, I., and Lim, Y. P. (2007). *The Brassica rapa Genome Sequencing Initiative*. Hong Kong: Genes, Genome and Genomics, Global Science Books 5, 35–39.

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 Bioinformatics* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Wang, H. Z. (2010). Review and future development of rapeseed industry in China. *Chin. J. Oil Crop Sci.* 32, 300–302. doi: 10.3724/SP.J.1011.2010.01385

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

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Wu, S., Han, B., and Jiao, Y. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. doi: 10.1016/j.molp.2019.10.012

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–268. doi: 10.1093/nar/gkm286

Yao, Y. X., Dong, Q. L., Zhai, H., You, C. X., and Hao, Y. J. (2011). The functions of an apple cytosolic malate dehydrogenase gene in growth and tolerance to cold and salt stresses. *Plant Physiol. Biochem. PPB* 49, 257–264. doi: 10.1016/j.plaphy.2010.12.009

Yoo, S. Y., Kim, Y., Kim, S. Y., Lee, J. S., and Ahn, J. H. (2007). Control of flowering time and cold response by a NAC-domain protein in Arabidopsis. *PloS ONE* 2, e642. doi: 10.1371/journal.pone.0000642

Yuan, Y., Chiu, L. W., and Li, L. (2009). Transcriptional regulation of anthocyanin biosynthesis in red cabbage. *Planta* 230, 1141–1153. doi: 10.1007/s00425-009-1013-4

Zeng, X., Xu, Y., Jiang, J., Zhang, F., Ma, L., Wu, D., et al. (2018). Identification of cold stress responsive microRNAs in two winter turnip rape (Brassica rapa L.) by high throughput sequencing. *BMC Plant Biol.* 18, 52. doi: 10.1186/s12870-018-1242-4

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

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., et al. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* 419, 779–781. doi: 10.1016/j.bbrc.2012.02.101

Zhang, Z., Zhang, Q., Wu, J., Zheng, X., Zheng, S., Sun, X., et al. (2013). Gene knockout study reveals that cytosolic ascorbate peroxidase 2(OsAPX2) plays a critical role in growth and reproduction in rice under drought, salt and cold stresses. *PloS ONE* 8, e57472. doi: 10.1371/journal.pone.0057472

Zhao, J., Paulo, M. J., Jamar, D., Lou, P., van Eeuwijk, F., Bonnema, G., et al. (2007). Association mapping of leaf traits, flowering time, and phytate content in Brassica rapa. *Genome* 50, 963–973. doi: 10.1139/G07-078

Zhou, D. M., Zhang, R. Z., Sun, W. C., Zhang, J., and Wang, H. L. (2014). Study on climatic suitability for winter rapeseed planting in arid and cold regions in North China. *Sci. Agric. Sin.* 47, 2541–2551.

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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fpls.2022.992391

OPEN ACCESS

EDITED BY
Ryo Fujimoto,
Kobe University,
Japan

REVIEWED BY
Shengwu Hu,
Northwest A&F University, China
Fu Shaohong,
Chengdu Agriculture and Forestry Academy of Sciences, China

*CORRESPONDENCE
Hui Feng
fenghuiaaa@syau.edu.cn

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

RECEIVED 12 July 2022
ACCEPTED 01 August 2022
PUBLISHED 18 August 2022

CITATION
Dong S, Zou J, Fang B, Zhao Y, Shi F, Song G, Huang S and Feng H (2022) Defect in *BrMS1*, a PHD-finger transcription factor, induces male sterility in ethyl methane sulfonate-mutagenized Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Front. Plant Sci.* 13:992391. doi: 10.3389/fpls.2022.992391

Defect in *BrMS1*, a PHD-finger transcription factor, induces male sterility in ethyl methane sulfonate-mutagenized Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*)

Shiyao Dong, Jiaqi Zou, Bing Fang, Ying Zhao, Fengyan Shi, Gengxing Song, Shengnan Huang and Hui Feng*

Department of Horticulture, Shenyang Agricultural University, Shenyang, China

Male sterility is an ideal character for the female parent in commercial hybrid seed production in Chinese cabbages. We identified three allele male sterile mutants *msm2-1*/2/3 in progenies of ethyl methane sulfonate mutagenized Chinese cabbage. It was proved that their male sterilities were controlled by a same recessive nuclear gene. Cytological observation showed that the delayed tapetal programmed cell death (PCD) as well as the abnormal pollen exine and intine led to pollen abortion in these mutants. MutMap combined with KASP analyses showed that *BraA10g019050.3C*, a homologous gene of *AtMS1* encoding a PHD-finger transcription factor and regulated pollen development, was the causal gene. A single-nucleotide mutation from G to A occurred at the 2443th base of *BrMS1* in *msm2-1* which results in premature termination of the PHD-finger protein translation; a single-nucleotide mutation from G to A existed at 1372th base in *msm2-2* that makes for frameshift mutation; a single-nucleotide mutation from G to A distributed at 1887th base in *msm2-3* which issues in the amino acid changed from Asp to Asn. The three allelic mutations in *BrMS1* all led to the male sterile phenotype, which revealed its function in stamen development. Quantitative reverse transcription polymerase chain reaction analysis indicated that *BrMS1* specially expressed in the anther at the early stage of pollen development and its expression level was higher in *msm2-1/2/3* than that in the wild-type "FT." BrMS1 was located at the nucleus and a length of 12 amino acid residues at the C-terminus had transcriptional activation activity. RNA-seq indicated that the mutation in *BrMS1* affected the transcript level of genes related to the tapetum PCD and pollen wall formation, which brought out the pollen abortion. These male sterile mutants we developed provided a novel gene resource for hybrid breeding in Chinese cabbage.

KEYWORDS

Chinese cabbage, male sterility, allelic mutants, PHD-finger transcription factor, RNA-seq

Introduction

Male sterility refers to the phenomenon in which stamens are degenerated or malformed, and viable male gametes are not produced, while female gametes are fertile in flowering plants. Male sterility has been widely used for hybrid seed production (Fan and Zhang, 2018). Up to now, many genic male sterility (GMS) genes have been identified in plants (Shi et al., 2015; Wan et al., 2019, 2020), and a great deal of effort has been made to develop male-sterility systems by using these GMS genes and their corresponding mutants to maintain and propagate GMS lines as female parents for hybrid seed production (Kim and Zhang, 2018). The male sterility system is also of immense significance in probing the molecular mechanism of pollen development. Spontaneous mutant male sterile mutants have been widely used in functional genomics (Liu et al., 2016; Liang et al., 2017; Zhou et al., 2017; Lin et al., 2020; Zhang et al., 2020). Due to its scarcity, artificial mutagenesis including physical mutagenesis, chemical mutagenesis, biological mutagenesis, and other methods to create male sterile mutants, has become an important measure for the study of gene regulatory networks for male gamete development (Ma et al., 2019; Tan et al., 2019; Huang et al., 2020).

Male gamete formation in anthers requires the coordinated participation of both sporophytic and gametophytic tissues (Hafidh et al., 2016). The tapetum, the innermost cell layer of the anther wall, plays a crucial role in regulating pollen development. The tapetum cells could provide the nutrients required for the meiosis of the microspore mother cell and the development of the microspore (Wu and Cheung, 2000; Parish and Li, 2010). During the tetrad period, the tapetum secrete sporopollenin precursor substances and they deposit continuously on the primexine to form a special T-shaped structure composed of the tectum, bacula, and foot layer (Shi et al., 2015). After the microspore mother cell completes meiosis, the tapetum synthesizes callose enzymes to degrade callose, which promotes the release of independent uninucleate microspores (Zhang et al., 2007; Parish and Li, 2010). Then, the tapetum undergoes programmed cell death (PCD), the *pollen coat proteins* (*PCPs*) and the tapetal cell debris produced by tapetum degradation, fill the pollen exine to form the pollen coat (Zhang et al., 2016). Therefore, abnormalities at any stage of the tapetum development could cause pollen aberrations.

The development of the tapetum is a precise and orderly process, which is jointly regulated by a variety of genes. To date, there are three major types of transcription factors (TFs) that regulate the development and function of the tapetum: *bHLH* TFs such as *DYT1* and *AMS*, *MYB* TFs such as *TDF1* and *MS188*, and *PHD-fingers* TFs such as *MS1* and *TIP3* (Zhu et al., 2008, 2011; Zhang and Yang, 2014; Wan et al., 2019; Yang et al., 2019a). These TFs cooperatively form the tapetal genetic pathway *(DYT1-TDF1-AMS-MS188-MS1)* and strictly control tapetum development (Zhu et al., 2011). Several orthologs of *MS1* have been identified in different species. In *Arabidopsis*, *AtMS1* directly regulated the expression of multiple *PCPs* such as *GRP14*, *GRP19* and affected the expression of *KCS7*, *KCS15*, and *KCS21* for *pollen coat lipids* (*PCLs*) synthesis (Lu et al., 2020; Zhang et al., 2021). In *rice*, *OsMS1/OsPTC1* can interact with *OsMADS15* and *TIP2*. It has been reported that *TIP2* coordinated with TDR to modulate the expression of *EAT1* and *Cys* protease gene (*CP1*) further regulated tapetal PCD and pollen exine formation in rice (Li et al., 2011; Yang et al., 2019b). In *maize, ZmMs7* can interact with maize *nuclear factor Y (NF-Y)* subunits to form *ZmMs7-NF-YA6-YB2-YC9/12/15* protein complexes that activate target gene *ZmMT2C* to regulate the tapetal development (An et al., 2020). However, the molecular mechanism by which the homologous gene of *MS1* in Chinese cabbage affects pollen development remains unclear.

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is one of the most important vegetable crops in eastern Asia, which is a typical cross-pollinated *Brassica* crops with obvious heterosis (Sharma et al., 2021). Breeding of male sterile lines is crucial for the commercialization of Chinese cabbage hybrid seed production (Singh et al., 2019). Mapping and cloning the male sterile genes are preconditions to the study of its molecular mechanism and applications. The genic male sterile lines could be created by artificial mutagenesis in Chinese cabbage. Tan et al. (2019) developed a male sterile mutant (*ftms*) from Chinese cabbage DH line "FT" by irradiating microspores with ^{60}Co γ-rays while *Bra010198* was a candidate gene for male sterile mutant (*ftms*), which encoded a putative *β-(1,3)*-galactosyltransferase that controls pollen exine development. Huang et al. (2020) obtained a stably inherited male sterile mutant (*msm*) by the identical methods and compared the differential expression genes of "FT" and *msm* flower buds using RNA-Seq technology.

Here, we identified three allele male sterile mutants named *msm2-1*/2/3 in an ethyl methane sulfonate (EMS)-mutagenized Chinese cabbage progenies. Cytological observation revealed delayed tapetum PCD, defective pollen exine and intine formation in *msm2-1* mutant. We discovered by MutMap and KASP that the non-synonymous base-pair mutations in *BrMS1* (*BraA10g019050.3C*) induced the pollen abortion in the allele mutants *msm2-1-2/3*. *BrMS1* was a homologous gene of *AtMS1*, which encoded a PHD-finger TF and regulated the pollen development. The protein encoded by the *BrMS1* was located at the nucleus in Chinese cabbage. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) performed that *BrMS1* expressed specifically in the anther at the early stage of pollen development and its expression level was higher in *msm2-1/2/3* versus in the wild-type "FT." The analysis of RNA-seq exposed that the expression level of genes associated with the tapetum PCD and pollen wall synthesis has changed significantly in the mutant *msm2-1*. Our research is the first to clone an *AtMS1* orthologs gene relevant to fertility regulation in Chinese cabbage, thereby providing basal information for applying this critical agronomic trait in hybrid breeding and studying the molecular mechanism of pollen development.

Materials and methods

Plant materials

The Chinese cabbage doubled haploid (DH) line "FT," obtained from a microspore culture of the commercial variety "Fukuda 50," was used as the wild type. We treated "FT" seeds with 0.8% EMS, the live plants (M_0) were self-pollinated, M_1 seeds were harvested. After M_1 self-pollination, we obtained stably inherited mutations in the M_2 generation (Gao et al., 2022). Among them, nine male sterile mutants with stable inheritance and fertility not affected by environment were identified after multiple generations phenotypic identification (Supplementary Figure S1). The fertile and sterile plants in AB lines of the nine male sterile materials were performed the allelic test by mutual cross. The materials whose segregation ratio of the hybrid progeny conformed to 1:1 were allelic mutants of each other (Supplementary Table S1). We screened out two groups of allelic mutant materials in total and select one group (*msm2-1*, *msm2-2*, and *msm2-3*) as the experimental materials for this study. All plant materials were cultured in a greenhouse at 20–28°C in Shenyang Agricultural University (Shenyang, P. R. China).

Genetic analysis

For the genetic analysis, the wild-type "FT" (P_1) was crossed with *msm2-1/2/3* mutants (P_2) to generate the F_1 and F_2 populations. Fertile plants and sterile plants from the F_2 population were selected for crossing to generate AB line. Phenotypic characterization was obtained for each generation (P_1, P_2, F_1, F_2 and AB line), and the segregation ratio of the population was analyzed by the Chi-square test.

Morphological observation of floral organs

To observe the difference in floral organs between the wild-type "FT" and *msm2-1/2/3*, the five floral organs (sepals, petals, filament, stamens, pistils) were collected at the full-bloom stage and dissected under a stereomicroscope (Nikon SMZ800, Japan). For investigating the pollen vitality, the fresh anthers at the full-bloom stage were extruded and immersed in 0.1% 2,3,5-triphenyltetrazolium chloride dye solution on a carrier plate covered with a coverslip. Pollen viability was analyzed *via* applying an optical microscope (Nikon ECLIPSE 80i, Japan).

Anther cytological observation

For understanding more about the period of pollen abortion, the flower buds of the wild-type "FT" and *msm2-1/2/3* were fixed using FAA solution (50% ethanol, 5% glacial acetic, 10% formalin), then dyed using safranin and fast green, and observed under an optical microscope (Nikon ECLIPSE 80i; Nikon, Japan). The buds were divided into six grades according to the length of the buds (BUD1 < 1.5 mm, BUD2: 1.5–2.0 mm, BUD3: 2.0–2.5 mm, BUD4: 2.5–3.0 mm, BUD5: 3.0–3.5 mm, BUD6 > 3.5 mm.). The detailed paraffin sections operation steps were performed according to the method described by Zhou et al. (2017).

For scanning electron microscopy (SEM), to observe the morphology of anther and pollen exine, the fresh anthers of the wild-type "FT" and the *msm2-1* mutant were collected at the full-bloom stage and fixed in 2.5% glutaraldehyde solution at 4°C for 48 h, processed according to the methods of Lin et al. (2014) and then examined under SEM (Hitachi TM3030, Japan).

For transmission electron microscopy (TEM), to further explore the morphology changes of pollen exine, the anthers are roughly divided into three periods: the early stage (BUDS < 2.0 mm), middle stage (2.0 mm < BUDS < 3.0 mm), and the late stage (BUDS > 3.0 mm) according to the pollen development process. Anthers at different pollen developmental stages from "FT" and *msm2-1* plants were examined by TEM (Hitachi Ltd., Tokyo, Japan) after a series of treatments as previously described (Wang et al., 2020).

Screening of the candidate genes by MutMap

The candidate gene was identified by the modified MutMap method. Fifty plants with mutant phenotype were selected from the F_2 plants, and a DNA pool was constructed by pooling equal amounts of plant leaf tissue from each plant. DNA was acquired from the wild-type "FT" and *msm2-1* mutant, the DNA pool was extracted using a DNA-secure Plant Kit (Tiangen, Beijing, China) and was re-sequenced using a NovaSeq 6,000 System sequencer (Illumina, San Diego, United States). The analysis method was the same as described by Gao et al. (2020) and Wang et al. (2020).

SNP genotyping through KASP

KASP was applied for the genotypic assay to detect the co-segregation of each single nucleotide polymorphism (SNP) identified by MutMap and confirmed the candidate gene. Allele-specific primers bearing the FAM and HEX fluorescence probes and the common primer were designed by LGC (Laboratory of the Government Chemist, Shanghai, China) and used to determine the male sterile genotypes of 183 F_2 plants (Supplementary Table S2). The primer mixed and PCRs were programmed as recommended by LGC. The KASP thermal cycling conditions were those described in Xi et al. (2018).

Fluorescence was detected with a QuantStudio 6 instrument (Applied Biosystems, Foster City, CA, United States).

Cloning and sequencing of the candidate gene

The full-length and CDS sequences of candidate genes were obtained by cloning and sequencing. Primers were designed according to the gene sequence information and are shown in Supplementary Tables S3, S4. We purified the PCR products with a Gel Extraction Kit (CWBIO, Beijing, China) and then introduced the purified products into a pGEM-T-Easy Vector (Promega, United States). The products were sequenced at GENEWIZ (Suzhou, China). We analyzed the sequences using DNAMAN V6 (Lynnon BioSoft, Canada).

Phylogenetic analysis and bioinformatic analysis of BrMS1 protein

The sequence and the information of BrMS1 were downloaded from the Brassica database. BrMS1 protein homologs in various plants were obtained from the GenBank database[1] and aligned with ClustalW (Larkin et al., 2007). A phylogenetic tree was constructed with MEGA6.0 by the neighbor-joining (NJ) method based on a bootstrap test of 1,000 replicates (Tamura et al., 2013). The conserved domain of BrMS1 was analyzed online at NCBI.[2]

Expression patterns analysis by qRT-PCR

Total RNA was isolated from different tissues of the wild-type "FT" and *msm2-1/2/3* at developmental stages with the TIANGEN RNA prep Pure Plant Kit. First-strand cDNA was synthesized using Fast Quant RT SuperMix (Tian gen). The cDNA was subsequently used as a template for qRT-PCR together with SYBR Green PCR Master Mix (TaKaRa, Dalian, China). ACTIN was used as the internal reference gene. Specific primers for each target gene are listed in Supplementary Table S5. Each sample had three biological replicates, each with three technical replicates.

Subcellular localization of BrMS1 proteins

To generate the *BrMS1–GFP* fusion protein, the full-length CDS of *BrMS1* was amplified. The coding regions of the *BrMS1* sequences were amplified without stop codons using the primers *BrMS1-GFP-F* and *BrMS1-GFP-R*. They were cloned into the C-terminal GFP fusion vector *pBWA(V)HS* driven by the *35S* promoter. Vector-bearing, *35S*-driven GFP was the negative control. Subcellular localization of the *BrMS1* proteins was performed according to Yoo et al. (2007). The constructs were then transiently transformed into tobacco (*Nicotiana benthamiana*) leaves through Agrobacterium tumefaciens infiltration. After 48 h, the epidermis of the tobacco leaves was injected with *DAPI* and examined with a confocal laser-scanning microscope (Leica Microsystems, Wetzlar, Germany). The primer sequences used in this experiment are shown in Supplementary Table S6.

Transactivation activity assay

The transcriptional activity assay was performed using the Matchmaker GAL4 Yeast Two-Hybrid System 3 (Clontech). To generate the BD–*BrMS1*, BD–ΔN, BD–ΔC and BD–PHD constructs, the full-length CDS and various deletions of *BrMS1* were PCR amplified and inserted into the pGBKT7 plasmid linearized. All recombinant plasmids and a pGBKT7 empty vector were separately transformed into the Y2H Gold Yeast strain. Then they were initially grown on the medium lacking Trp for 3 days at 30°C. Four of the resulting independent colonies were transferred onto the medium without Trp, His, and Ade for 3 days at 30°C to assay transcriptional activity. The primers used for vector construction are listed in Supplementary Table S7.

RNA-seq

To identify transcripts involved in the regulation of pollen abortion in the Chinese cabbage, we performed an extensive transcriptomic analysis of entire anther. Total RNA was extracted from the wild-type "FT" and *msm2-1* with the RNA prep Pure Plant Plus Kit (Tiangen, Beijing, China). Samples were collected with three biological replicates. Six cDNA libraries (F1, F2, F3, S1, S2, and S3) were constructed and sequenced using the Illumina HiSeq™ 2,500 sequencing platform at Beijing Genomics of Institute (BGI), Shenzhen, China. The detailed method and procedure have been described previously by Huang et al. (2017). In this study, different expression genes (DEGs) were defined as those genes with a false discovery rate (FDR) ≤ 0.01 and an absolute log2 ratio value ≥ 10. To further study the biological functions and metabolic pathways of the DEGs, we performed Gene Ontology (GO)[3] functional analysis and Kyoto Encyclopedia

1 http://www.ncbi.nlm.nih.gov/BLAST

2 https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

3 http://www.geneontology.org/

of Genes and Genomes (KEGG) pathway enrichment analysis of these genes (Ashburner et al., 2000; Kanehisa et al., 2007).

Results

Genetic analysis and morphological characteristics of the male sterile mutants *msm2-1/2/3*

To elucidate the genetics of the male sterile phenotype, the wild-type "FT" (P_1) and the mutants *msm2-1/2/3* (P_2) were employed as the male parent and the female parent to generate F_1 and F_2 populations. All F_1 individuals displayed male fertile phenotype. As disclosed by the Chi-square test, the segregation rate of the F_2 population corresponded to the expected Mendelian ratio of 3:1 (fertile:sterile; Table 1). Allelism test showed that the fertility separation ratio of *msm2-1*, *msm2-2* and *msm2-3* hybrid offspring was 1:1 (fertile:sterile), which implied that mutant *msm2-1*, *msm2-2* and *msm2-3* were controlled by an allelic gene (Table 2). Therefore, these results indicated that the male sterile trait of *msm2-1/2/3* mutants was controlled by a single nuclear recessive gene. From the perspective of breeding, it also demonstrated that allele gene mutation result in male sterility in the mutants *msm2-1/2/3*.

At the full-bloom stage, the morphology of floral organs was investigated. The stamens of *msm2-1/2/3* had developed normally yet no pollen attached to anthers (Figure 1A). Other flower organs in the *msm2-1/2/3* mutants were found to be slightly smaller than those in the wild-type "FT" (Figures 1B–G). Pollen viability observations also illustrated that the wild-type "FT" could generate mature and active pollen grains, whereas *msm2-1/2/3* could not produce pollen grains (Figures 1H–K). These three *msm2-1/2/3* mutants were allelic as evidenced by allelic test and manifested similar phenotypes under the same growth conditions; thus, some of the subsequent analyses in this study were only performed on *msm2-1*.

TABLE 1 Analysis for the genetic characteristics of the mutants *msm2-1/2/3*.

Generation	Total	Male ferility	Male sterility	Segregation radio	χ^2
P_1("FT")	1	1	0		
P_2(*msm2-1*)	1	0	1		
F_1(P_2×P_1)	100	100	0		
F_2	314	239	75	3.19:1	0.21
P_2(*msm2-2*)	1	0	1		
F_1(P_2×P_1)	100	100	0		
F_2	122	96	26	3.69:1	0.89
P_2(*msm2-3*)	1	0	1		
F_1(P_2×P_1)	100	100	0		
F_2	63	49	14	3.50:1	0.26

Cytological observation of the anther development

To explore the cytological characteristics of pollen abortion in *msm2-1/2/3*, we examined paraffin sections prepared at different anther development stages of the wild-type "FT" and the *msm2-1/2/3*. The paraffin section analysis showed there was no significant difference in *msm2-1/2/3* compared with the wild-type "FT" at the early stage of the anther development (Figures 2A–H). At the middle stage of anthers development, the wild-type tapetum was concentrated with less vacuolation and darker staining. However, the tapetum cells became more vacuolated and less stained in *msm2-1/2/3* (Figures 2I–L). Then, the wild-type tapetum cell appeared to undergo degeneration, and the cytoplasmic constituents remained densely stained. By contrast, we discovered that the tapetum PCD was retarded in the *msm2-1/2/3* mutants and the tapetum cells had vacuolated (Figures 2M–T). At the late stage of anthers development, the tapetal cells degenerated and the pollen grains matured with a round shape in the wild-type "FT." Nevertheless, cell debris of tapetal cells and immature pollen grains remained in the *msm2-1/2/3* (Figures 2U–X). Paraffin sections of anther indicated that the degradation of tapetum cell delayed occurred in mutants *msm2-1/2/3*.

To investigate the anther and pollen exine morphology, the fresh anthers of the wild-type "FT" and the *msm2-1* were inspected by scanning electron microscopy. The SEM observation of the anther and pollen surface depicted the *msm2-1* mutant had a slightly smaller anther versus the wild type at the anther maturation stage (Figures 3A,E). The cuticle ridges of the anther epidermis in wild-type "FT" are loosely distributed while the cuticles of *msm2-1* are arranged relatively tightly (Figures 3B,F). The mature pollen grains of wild-type "FT" rounded surfaces having a reticular structure and the visible germinal furrow (Figures 3C,D), while pollen grain surface of mutant *msm2-1* is smooth with neither reticular structure nor germination furrow, and finally degraded to pieces (Figures 3G,H).

To reveal the *msm2-1* pollen grain's developmental defects in detail, TEM was conducted to observe the anther development. In the early stage of the anther development, there was no distinct difference between the wild-type "FT" and the *msm2-1* mutant (Figures 4A,D,G,J). In the middle stage of the anther development, the tapetum cells of wild-type "FT" began to degrade and secrete lipid droplets and protein substances (Figure 4H). These substances were deposited on the wild-type pollen exine, which formed with the well-organized layers including the tectum, bacula, and foot layer (Figure 4B). The degradation of the tapetum was delayed in *msm2-1* mutant (Figure 4K) and the skeletal structure of the pollen exine was thickened continuously (Figure 4E). At the late stage of the anther development, the pollen coat substances produced by

tapetum degradation filled the sexine of the pollen exine to form the pollen coat in the wild-type "FT" (Figures 4C,I). However, the pollen of *msm2-1* was surrounded by a thick abnormal pollen exine structure, while the filling of skeleton structure by pollen coats was not observed in *msm2-1* (Figure 4F). In the wild type, a white pollen intine structure could be seen on the inner side of the pollen wall, while the pollen inner wall structure of the mutant *msm2-1* was vague and amorphous (Figures 4C,F).

Identification of the causal gene of *msm2-1* by MutMap and KASP

A modified MutMap method was applied to identify the mutant gene. By genome resequencing of the wild-type "FT," *msm2-1* mutant and the F_2 mutant DNA pools, 93,975,658, 58,702,874 and 166,730,557 high-quality reads were obtained, respectively. Whereafter, these clean reads were aligned with the reference genome, resulting in 98.00%, 99.00% and 98.83% mapped to it, respectively. With the SNP index of 0.95 as the threshold, we mapped 1.08 Mb (14,034,596–15,116,841) candidate regions on chromosome A10 and detected three SNP mutations (SNP A10, 140,813,37; SNP A10, 143,901,24; SNP A10, 146,585,70), which were located at exons and resulted in nonsynonymous mutations (Figure 5A; Supplementary Table S8).

For further determining the candidate genes, we adopted KASP technology to design primers for three candidate SNPs, and genotyping was performed in "FT," mutant *msm2-1*, F_1 and F_2 mutant plants. The results showed that the genotypes of three SNPs in wild type, mutant and F_1 generation were G: G, A: A and G: A, respectively. The gene phenotypes of *BraA10g019050.3C* SNP (A10:14,081,337) in F_2 generation were all A: A (Figure 5B) and the SNPs of the other two genes (A10:14,390,124; A10:14,658,570) appeared six and nine heterozygous types (G: A) in the F_2 generation sterile plants (Supplementary Figure S2). Therefore, *BraA10g019050.3C* was co-segregated with the mutant phenotype, which further confirmed that *BraA10g019050.3C* was the mutant gene of *msm2*-1. *BraA10g019050.3C* encodes a PHD-finger protein sharing 88% identity with the *Arabidopsis MS1*, which functions as a transcriptional activator that regulates tapetal development and pollen exine biosynthesis. Hence, *BraA10g019050.3C* was named *BrMS1*.

TABLE 2 Analysis for allelism test of the mutants *msm2-1/2/3*.

Generation	Total	Male ferility	Male sterility	Segregation radio	χ^2
msm2-1×*msm2-2*	50	23	27	0.85:1	0.32
msm2-1×*msm2-3*	50	26	24	1.08:1	0.08
msm2-2×*msm2-3*	49	25	24	1.04:1	0.02

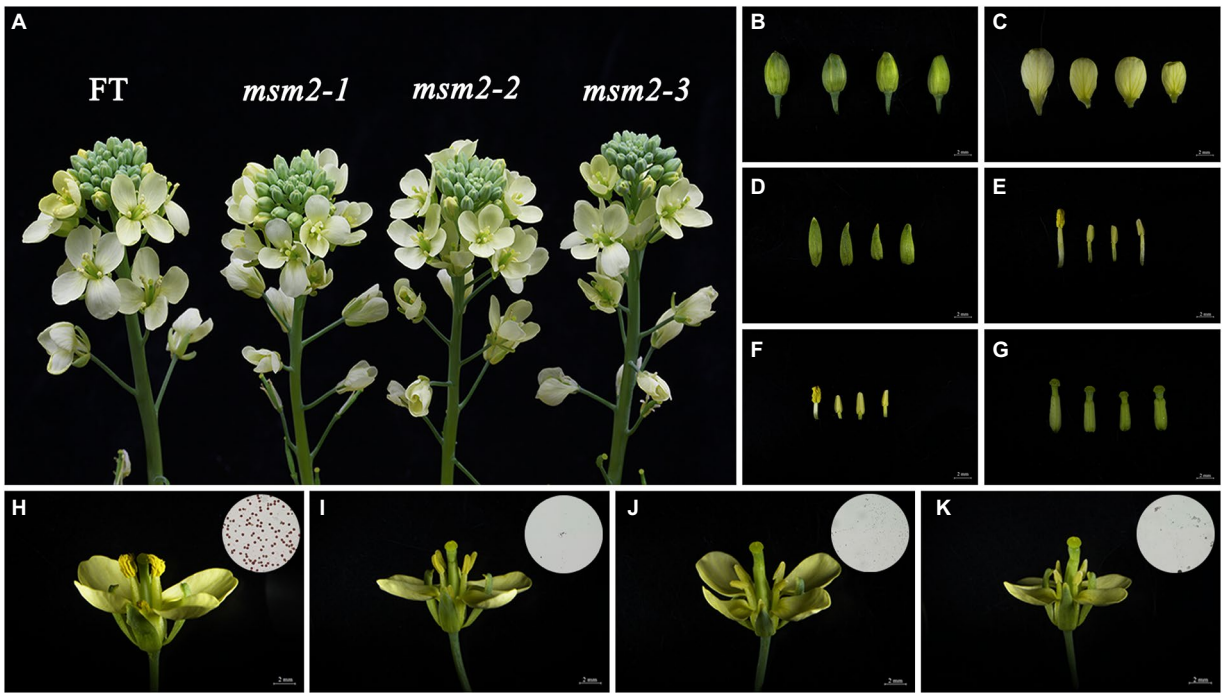


FIGURE 1
Phenotypic analysis of the wild-type "FT" and *msm2-1/2/3* mutants. Comparison of the floral organs of the wild-type "FT" and mutant *msm2-1/2/3* (from left to right). **(A)** Inflorescence **(B)** buds, **(C)** petals, **(D)** sepals, **(E)** stamens (long), **(F)** stamens (short), **(J)** pistils. Scale bar 2 mm analysis of pollen viability of the wild-type "FT" **(H)**, *msm2-1* **(I)**, *msm2-2* **(J)**, and *msm2-3* **(K)**. Scale bar 150 μm.

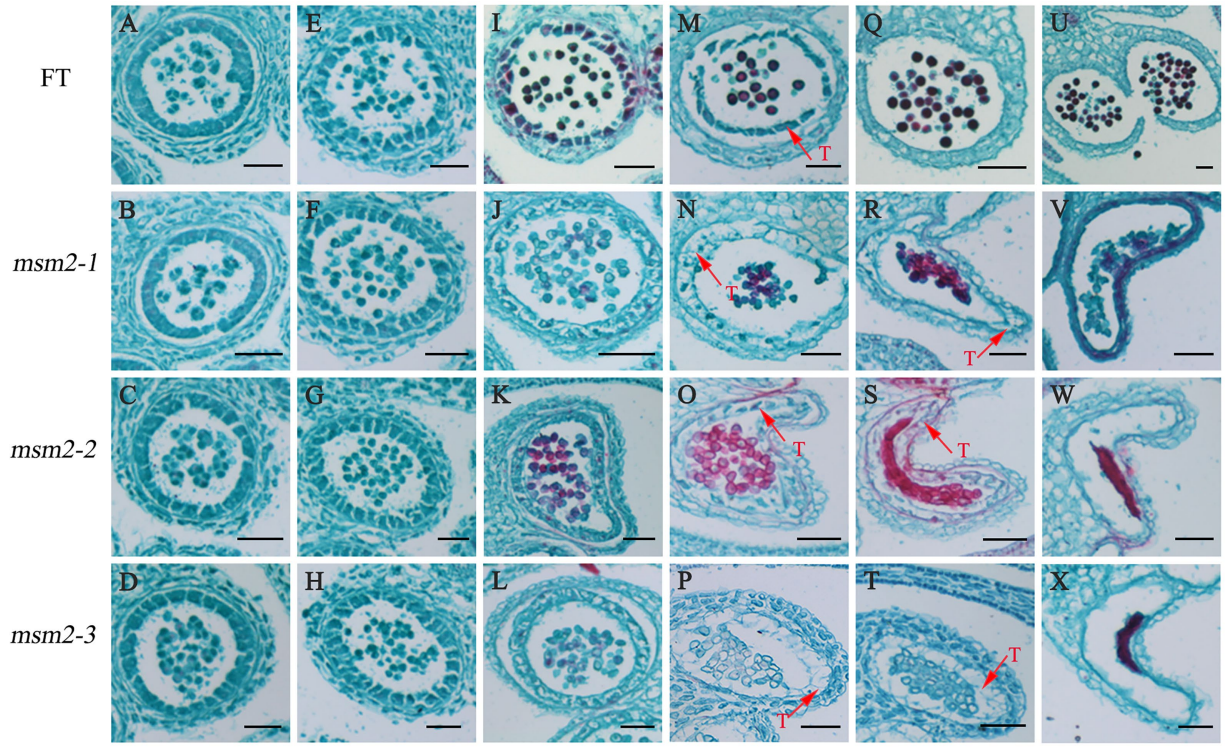


FIGURE 2
Anther paraffin section observation of the wild-type "FT" and *msm2-1/2/3* mutants. **(A–D)** The stage of tapetal formation (i.e., the tetrad period), BUD1<1.5 mm; **(E–H)** The stage of the tapetum cell structure integrity and clearly demarcated (i.e., the uninucleate microspore stage), BUD2: 1.5–2.0 mm; **(I–L)** The stage of tapetal cytoplasm dense, BUD3: 2.0–2.5 mm; **(M–P)** The stage of tapetum cells begin the degradation, BUD4: 2.5–3.0 mm; **(Q–T)** The stage of tapetum cells complete the degradation, BUD5: 3.0–3.5 mm; **(U–X)** The stage of anthers dehiscence and mature pollen release, BUD6>3.5 mm. T tapetum. Scale bar, 50 μm.

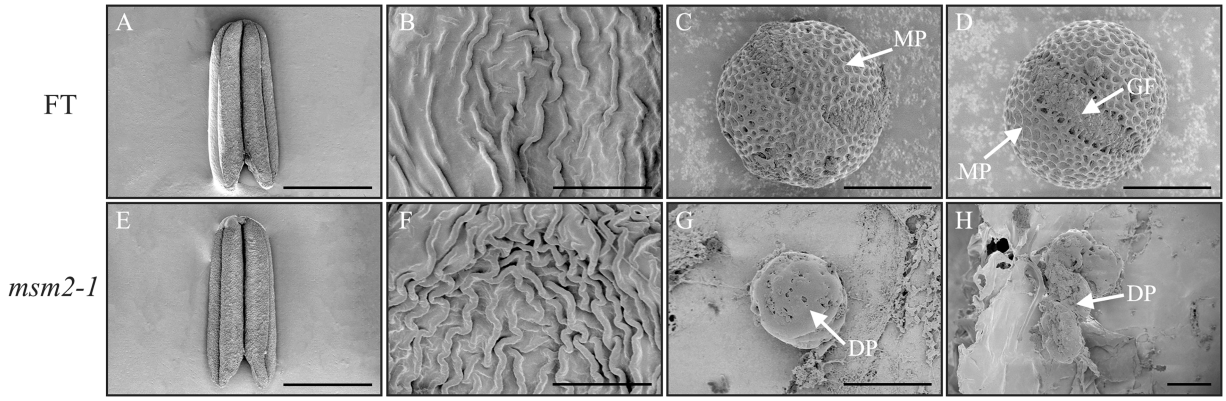


FIGURE 3
SEM analysis of the anthers and pollen grains in the wild-type "FT" and *msm2-1* mutant. The anthers of the wild-type "FT" **(A)** and *msm2-1* mutant **(E)**. Scale bar, 1 mm. The anthers epidermis of the wild-type "FT" **(B)** and *msm2-1* mutant **(F)**, Scale bar. 5 μm. The surface on the pollen exine of the wild-type "FT" **(C,D)** and *msm2-1* mutant **(G,H)**. *MP* mature pollen grains. *GF* germinal furrow, *DP* degenerated pollen grains. Scale bar, 10 μm.

Sequence analysis of *BrMS1* in *msm2-1/2/3* mutants

The mutants *msm2-1*, *msm2-2* and *msm2-3* are reciprocal allelic mutants, which means that the function of *BrMS1* can be verified in the other two mutants. The results of the *BraA10g019050.3C* gene cloned from three mutants proved that at 2443 bp in exon 3 of the mutant *msm2-1*, the base sequence changed from G to A and the amino acid changed from tryptophan to terminator codon, leading to the

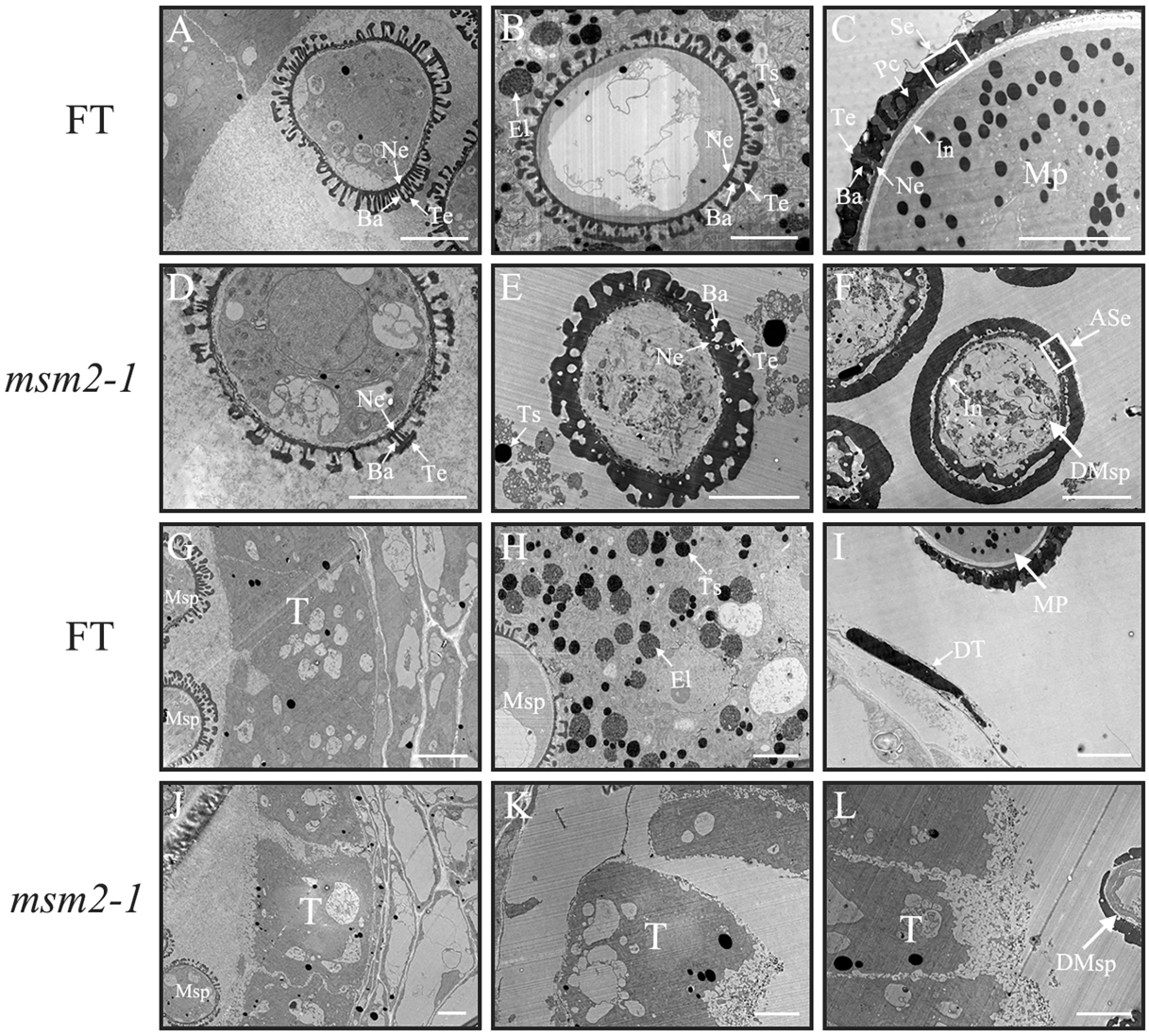


FIGURE 4
TEM analysis of pollen wall and tapetum development in the wild-type "FT" and *msm2-1* mutant. The process of pollen exine development in wild-type "FT" and *msm2-1* mutant at early (BUDS<2.0 mm; **A,D**), middle (2.0 mm<BUDS<3.0 mm; **B,E**), late (BUDS>3.0 mm; **C,F**) stage. The process of tapetum development in wild-type "FT" and *msm2-1* mutant at early (BUDS<2.0 mm; **G,J**), middle (2.0 mm<BUDS <3.0 mm; **H,K**), late (BUDS>3.0 mm; **I,L**) stage; Scale bar, 5 μm. T, Tapetal layer; DT, Degraded tapetum; Msp, Microspores; DMsp, Degenerated microspores; MP, Mature pollen grains, Ba, Bacula; Te, Tectum; Se, Sexine; In, Intine; ASe, Abnormal sexine; El, Elaioplast; Ts, Tapetosome.

termination of translation; At 1372 bp in the end of the second intron of mutant *msm2-2*, the base sequence changed from G to A, resulting in cutting off by mistake at the first base of third exon and giving rise to frameshift mutation; At 1887 bp in the third exon of *msm2-3*, the base sequence changed from G to A, and the amino acid changed from Asp to Asn (Figures 6A,B). The amino acid transformation attributed to the *BrMS1* mutation likewise contributed to differences in the three-dimensional structure of the protein (Figures 6C–F). These results hinted that the mutation in *BrMS1* led to the male sterile phenotype in *msm2-1/2/3* and verified its function in stamen development.

Structural and phylogenetic analyses of BrMS1

To elucidate the evolutionary relationship between BrMS1 and its close homologs, an unrooted tree of BrMS1 and its 16 homologs from moss to flowering plants was constructed (Figure 7A). BrMS1 was clustered with many dicotyledons, and located in the same clade as AtMS1, BnMS1, BoMS1. BrMS1 has a zinc finger protein (PHD-finger) domain (Figure 7B). The PHD-finger domain is functionally conserved in a variety of plants. We compared the sequences of the PHD-finger structural regions among BrMS1 and homologous sequences in *Arabidopsis*, *rice*,

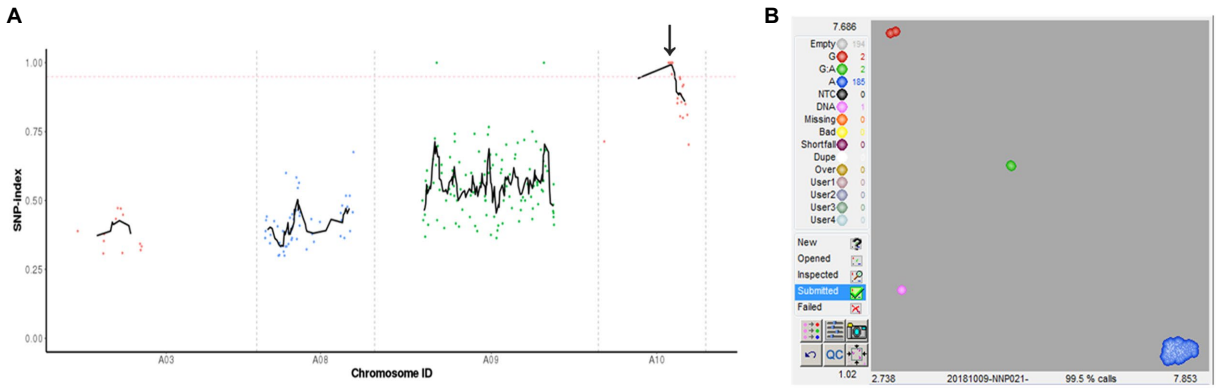


FIGURE 5
Identification of the *msm2-1/2/3* candidate gene. **(A)** SNP index plot of ten chromosomes was produced by MutMap analysis. The x-axis represents the positions of the ten chromosomes and the y-axis represents the SNP index. The dotted pink line represents the index threshold (0.95). **(B)** The results of SNP genotyping arrays of *BraA10g019050.3C*.

maize, and *wheat*. The amino acid sequence of the PHD-finger domain sequence in Chinese cabbage was 90% consistent with that of AtMS1 (Figure 7C).

Spatio-temporal expression pattern of *BrMS1* in *msm2-1/2/3* mutants

qRT-PCR analysis was conducted to recognize the spatial and temporal expression patterns of *BrMS1* in wild-type "FT" and mutants *msm2-1/2/3*. The expression level of *BrMS1* in the buds of the mutants *msm2-1/2/3* was higher than in the wild-type "FT" line and the expression level of the *BrMS1* gene in the mutant *msm2-1/2* was significantly higher than that of the wild-type "FT," while the expression level of the *BrMS1* gene in *msm2-3* was slightly higher than that of the wild-type "FT" (Figure 8A). The results were likewise made clear that *BrMS1* was specifically expressed in stamens at stage I and stage II (i.e., the stage of tetrad to uninucleate microspore release). Afterwards, the expression level of *BrMS1* declined rapidly and was almost undetectable in the mature anthers (Figures 8B,C). The expression patterns of *BrMS1* in allelic mutants *msm2-2/3* are same as that in mutant *msm2-1* (Supplementary Figure S3).

BrMS1 functions as a transcription factor with transcriptional activation activity

Previous studies have shown that *Arabidopsis MS1* functions as a transcriptional activator that regulates pollen and tapetum development (Ito and Shinozaki, 2002). Herein, to verify if *BrMS1* functions as a TF in Chinese cabbage, a subcellular localization assay was initially performed. The transient expression assay showed that the GFP signal was only detected in the protoplast nucleus and indicated that *BrMS1* encodes a PHD-finger protein localized in the nucleus, supporting that it functions as a TF (Figure 9A).

For analyzing whether *BrMS1* exhibits transcriptional activity activation, the full-length CDS and three truncated fragments of *BrMS1* (ΔN, ΔC, PHD) were ligated into a pGBKT7 vector to generate BD-*BrMS1*, BD-ΔN, BD-ΔC, and BD-PHD fusion constructs (Figure 9B). The results showed all of vectors transformed yeast cells could grow in the control medium (SD/-Trp), but only BD-*BrMS1* and BD-ΔC could grow in the selective medium (SD/-Trp/-His/-Ade). The findings demonstrated that *BrMS1* exhibited transcriptional activation activity in yeast (Figure 9C). Subsequently, the characteristics of the deleted fragments were extensively compared. We discovered that a length of 12 C-terminal amino acid residues were required to activate activity. Taken together, these results suggested that *BrMS1* was a nuclear-localized TF with a length of 12 C-terminal amino acid transcriptional activation activity.

BrMS1 affects the transcription level of genes associated with tapetal PCD and pollen wall development

As a TF, *BrMS1* is expected to affect the anther development by regulating the expression of downstream genes. To identify genes that might be regulated by *BrMS1*, we performed a comparative transcriptome analysis of anther in wild-type "FT" and the *msm2-1* using RNA-Seq. 4,779 DEGs were identified between the wild-type "FT" and the *msm2-1*, involving 1,475 upregulated genes and 3,304 downregulated genes in *msm2-1* (fold change ≥ 10, false discovery rate, FDR < 0.01; Figure 10A). The KEGG analysis displayed that DEGs were enriched in multiple metabolic pathways associated with pollen development, including the pentose and glucuronate interconversions pathway (53 DEGs), phenylpropanoid metabolism pathway (52 DEGs), galactose metabolism pathway (21 DEGs) as well as cutin, suberine and wax biosynthesis pathway (15 DEGs;

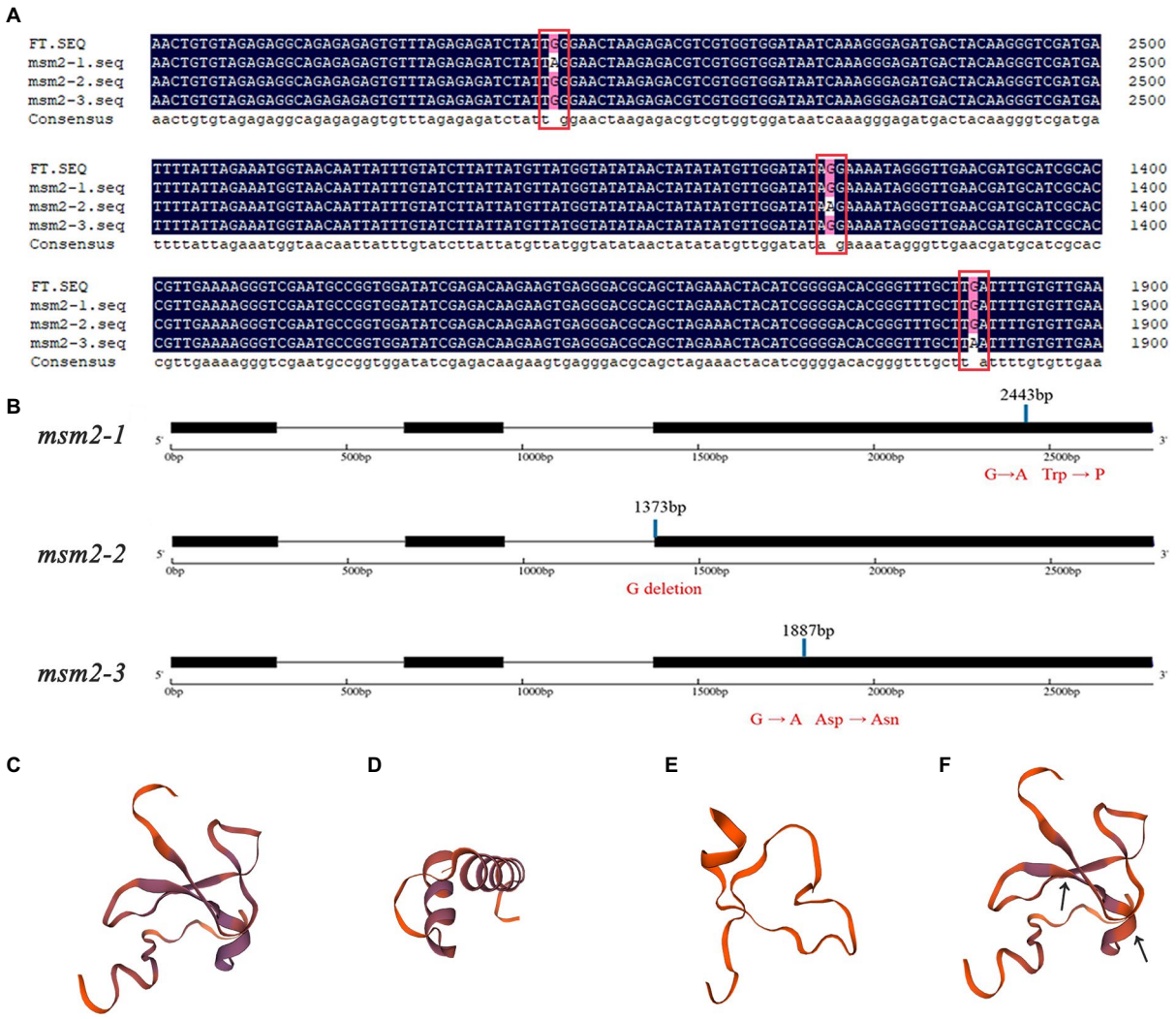


FIGURE 6
Sequence alignment and structure analysis of *BrMS1* between the wild-type "FT" and *msm2-1/2/3* mutants. Comparison of the *BraA10g019050.3C* nucleotide sequence **(A)** and structure **(B)** between the wild-type "FT" and *msm2-1/2/3* mutants. Three-dimensional protein structure of *BraA10g019050.3C* in the wild-type "FT" **(C)**, *msm2-1* **(D)**, *msm2-2* **(E)** and *msm2-3* **(F)**.

Figure 10B). In the pentose and glucuronate interconversions pathway, several genes related to pectin biosynthesis which affected the formation of pollen intines were identified, containing 10 pectinesterase inhibitor genes (*PMEIs*), 13 polygalacturonases genes (*PGs*) and 11 pectate lyasegenes (*PLs*; Figure 10C; Supplementary Table S9). Most of them were down-regulated in *msm2-1* and some of these genes expressed specially in the wild-type "FT" such as *PMEIs* (*BraA06g033760.3C* and *BraA08g011580.3C*), *PGs* (*BraA02g039370.3C*, *BraA02g039400.3C* and *BraA09g065620.3C*), *PLs* (*BraA09g059390.3C* and *BraA06g039690.3C*). The phenylpropanoid pathway splits into two main branches leading to hydroxycinnamic acids and flavonoids, both of which are related to the synthesis of pollen exine (Preston et al., 2004). The expression levels of some genes involved in the synthesis of key enzymes e.g., cinnamyl alcohol dehydrogenase, feruloyl-CoA, caffeoyl-CoA were significantly changed in the mutant *msm2-1*, among which *COMT* (*BraA02g024180.3C*) was specifically expressed in the wild-type "FT" (Figure 10D; Supplementary Table S10). RNA-seq profiling suggested that the defect in *BrMS1* affected the transcript levels of genes involved in pollen intine and exine biosynthesis metabolic pathways in *msm2-1*.

In order to assess whether the defect in *BrMS1* broadly affects other pollen development pathways, we compared gene expression associated with pollen development between the wild-type "FT" and the *msm2-1*. In *Arabidopsis*, *AtMS1* regulates the expression of *PCPs* and *PCLs* genes, which participate in pollen coat formation (Lu et al., 2020; Zhang et al., 2021). In our study, several *PCPs* genes such as *GRPs* and *EXLs* exhibited a prominent reduction in the mutant *msm2-1*, but no significant changes were observed in the expression of *PCL* synthesis genes such as *KCS7/5/11/21* (Supplementary Table S11). Excessive deposition of sporopollenin on the pollen primexine of the *msm2-1* was observed by TEM. The

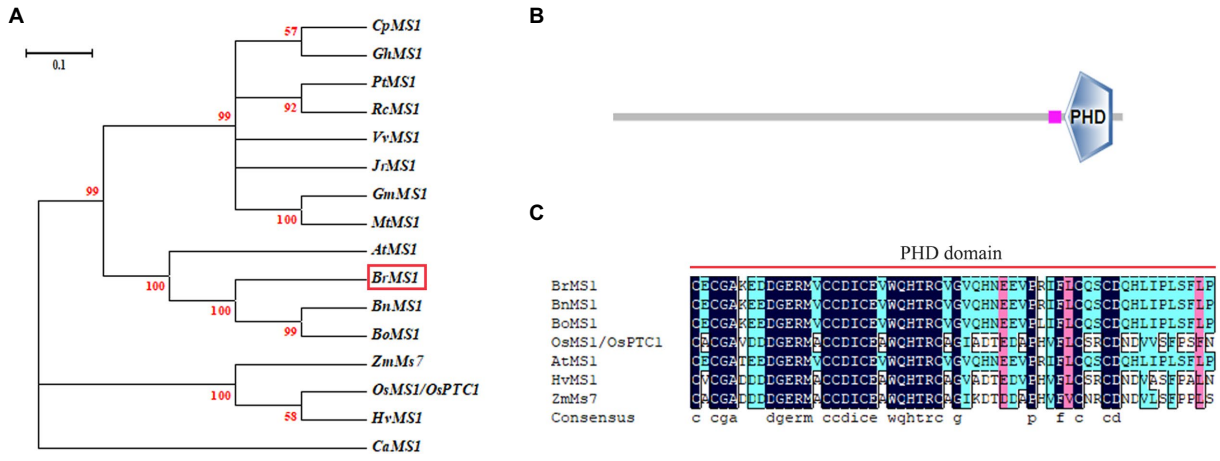


FIGURE 7
PHD-finger domain sequence alignment in different species and phylogenetic analysis of *MS1*. **(A)** Phylogenetic analysis of MS1 and its related proteins. An unrooted maximum likelihood tree was created by MEGA 3.1 using the MS1-related sequences from *Carica papaya* (*Cp*), *Gossypium hirsutum* (*Gh*), *Populus trichocarpa* (*Pt*), *Ricinus communis* (*Rc*), *Vitis vinifera* (*Vv*), *Juglans regia* (*Jr*), *Glycine max* (*Gm*), *Medicago truncatula* (*Mt*), *Arabidopsis* (*At*), *Brassica rapa* (*Br*), *Brassica napu* (*Bn*), *Brassica oleracea* (*Bo*), *Zea mays* (*Zm*), *rice* (*Os*), *barley* (Hv), *Capsicum annuum* (*Ca*). **(B)** Conserved domain structure of *BrMS1*. **(C)** PHD-finger domain sequence alignment. Residues with 100% similarity are shaded in black. Those with ≥75% similarity are shaded in pink. Those with ≥50% similarity are shaded in blue.

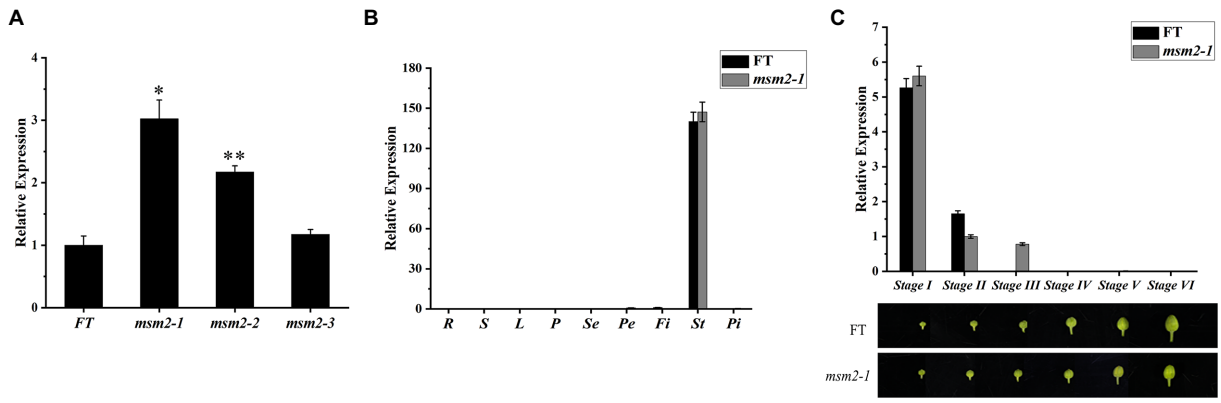


FIGURE 8
Expression pattern of *BrMS1* between the wild-type "FT" and *msm2-1/2/3* mutants. **(A)** Analysis of *BrMS1* expression in the wild-type "FT" and the allelic mutants *msm2-1/2/3*. Asterisks indicate statistically significant differences determined using a *t*-test (*p<0.05; **p<0.01). **(B)** Expression of *BrMS1* in the roots (*R*), stems (*S*), leaves (*L*), pods (*P*), sepal (*Se*), petal (*Pe*), filament (*Fi*), stamen (*St*) and pistil (*Pi*) of the wild-type "FT" and *msm2-1* mutant. **(C)** Analysis of *BrMS1* expression in different grades of buds in wild-type "FT" and *msm2-1* mutant. The grading rules for flower buds are the same as those for paraffin sections.

results of the RNA-seq also showed that some crucial sporopollenin synthesis genes (*SSGs*) such as *ACOS5*, *MS2*, *CYP703A2*, *PKSA* were up-regulated in the *msm2-1* (Supplementary Table S12). In the tapetum development regulatory network, *MS188* is the upstream regulatory factor of *MS1*. The expression of *MS188* genes (*BraA10g013970.3C* and *BraA02g013010.3C*) in the mutant *msm2-1*, was higher than in the wild-type "FT." *AMS* (*BraA07g004220.3C* and *BraA03g043400.3C*), which was the upstream gene of *MS188*, also up-regulated expression in *msm2-1* (Supplementary Table S13). The degradation of tapetum cells is regarded as a process of PCD, and papain-like cysteine proteases play a key control role in the timely degradation of tapetum cells (Liu et al., 2018; Wang et al., 2021). *CEP1* is a crucial executor during tapetal PCD and that proper *CEP1* expression is necessary for timely degeneration of tapetal cells and functional pollen formation in *Arabidopsis thaliana* (Zhang et al., 2014). *BraA03g026210.3C*, a homologous gene of *CEP1*, was up-regulated in *msm2-1*. We also observed the downregulation of the other two cysteine proteases genes *CP1* (*BraA10g004970.3C* and *BraA08g033980.3C*) in *msm2-1* compared with the wild-type "FT" (Supplementary Table S13). Overall, these

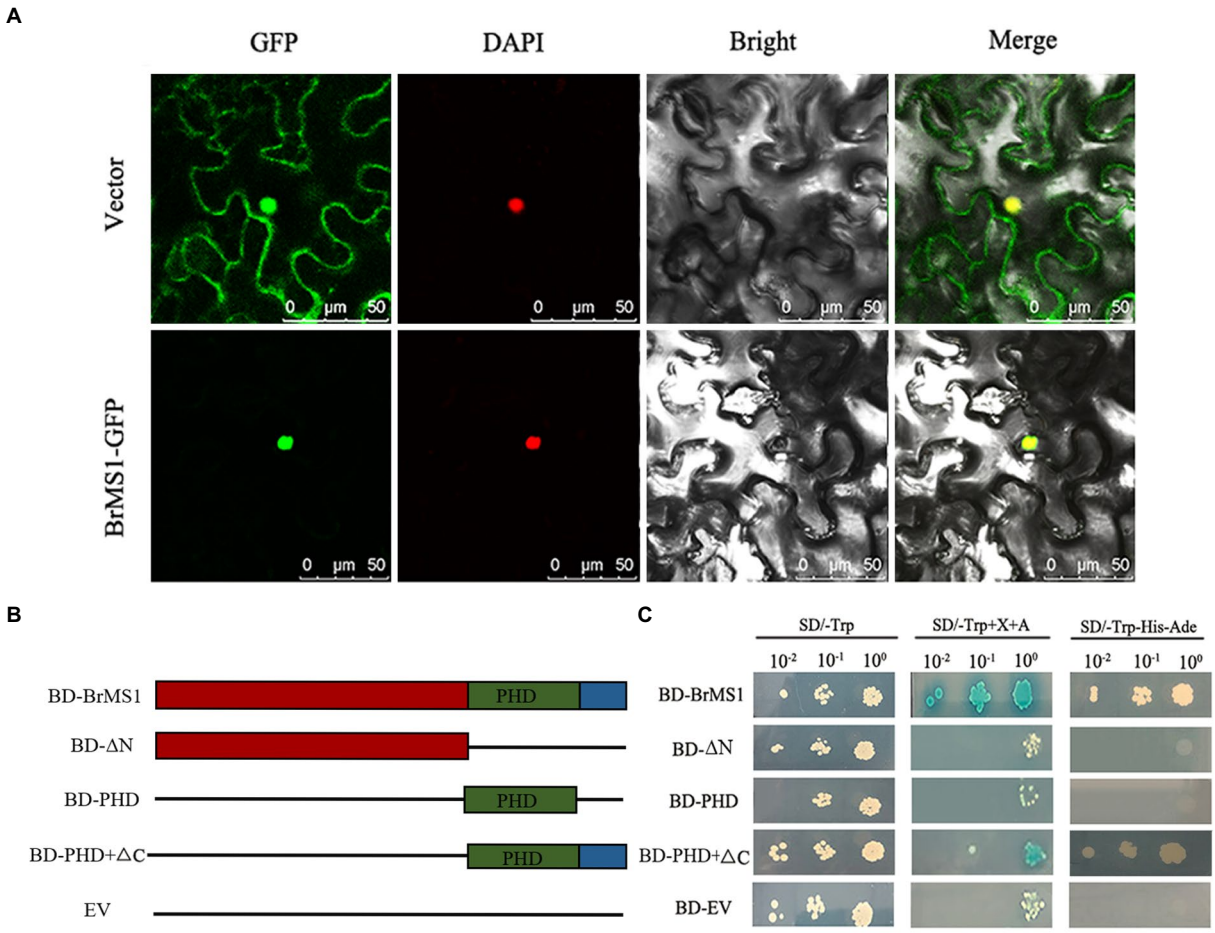


FIGURE 9
Subcellular localization and transcriptional activation assay of the *BrMS1*. **(A)** Subcellular localization of *BrMS1* in Tobacco leaf. DAPI staining was used as a nuclear marker. Scale bar, 50 μm. **(B)** The schematic diagram displays various truncated constructs and pGBKT7 empty vector used for the transcriptional activation assay. **(C)** Transcriptional activation assay of *BrMS1* in the Y2H yeast strain. BD GAL4–DNA binding domain, EV empty vector. The BD–EV used as a negative control.

findings suggested that *BrMS1* might directly or indirectly regulate the expression levels of *PCPs*, *SSGs* and *Cys* proteases genes, further resulting in pollen abortion in *msm2-1*.

To test the reliability of RNA-seq data, we verified the transcript levels of four PCP genes and three cysteine protease genes between the wild-type "FT" and the mutant *msm2-1* by qRT-PCR, and the analyses showed that their mRNA levels were all significantly reduced in *msm2-1*, apart from *CEP1*, which was up-regulated (Figure 11). These findings are consistent with the results from the RNA-seq analyses.

Discussion

The application of male sterility lines is an effective approach to commercial hybrid seed production and heterosis utilization in Chinese cabbage. Moreover, male sterility mutant is an important material to study pollen development and nuclear-cytoplasmic interactions. The creation of male sterile materials through artificial mutation has become a crucial mean of male sterility research. Discovery and research on male sterile mutants also had been a hot topic in research in various species. In this study, we cloned firstly the male sterility gene *BrMS1* in Chinese cabbage by forwarding genetics strategy and revealed the critical role of *BrMS1* in pollen development. In addition, the male sterile mutants we discovered have the characteristics of complete abortion, genetical stability and no concomitant adverse traits, which can be adopted as excellent sources of male sterility for Chinese cabbage hybrid seed production.

The defect of *BrMS1* caused tapetum cell delayed degeneration and abnormal pollen wall in *msm2-1/2/3* mutants

Tapetum timely degradation is essential for the pollen development in plants. The forms of abnormal tapetal

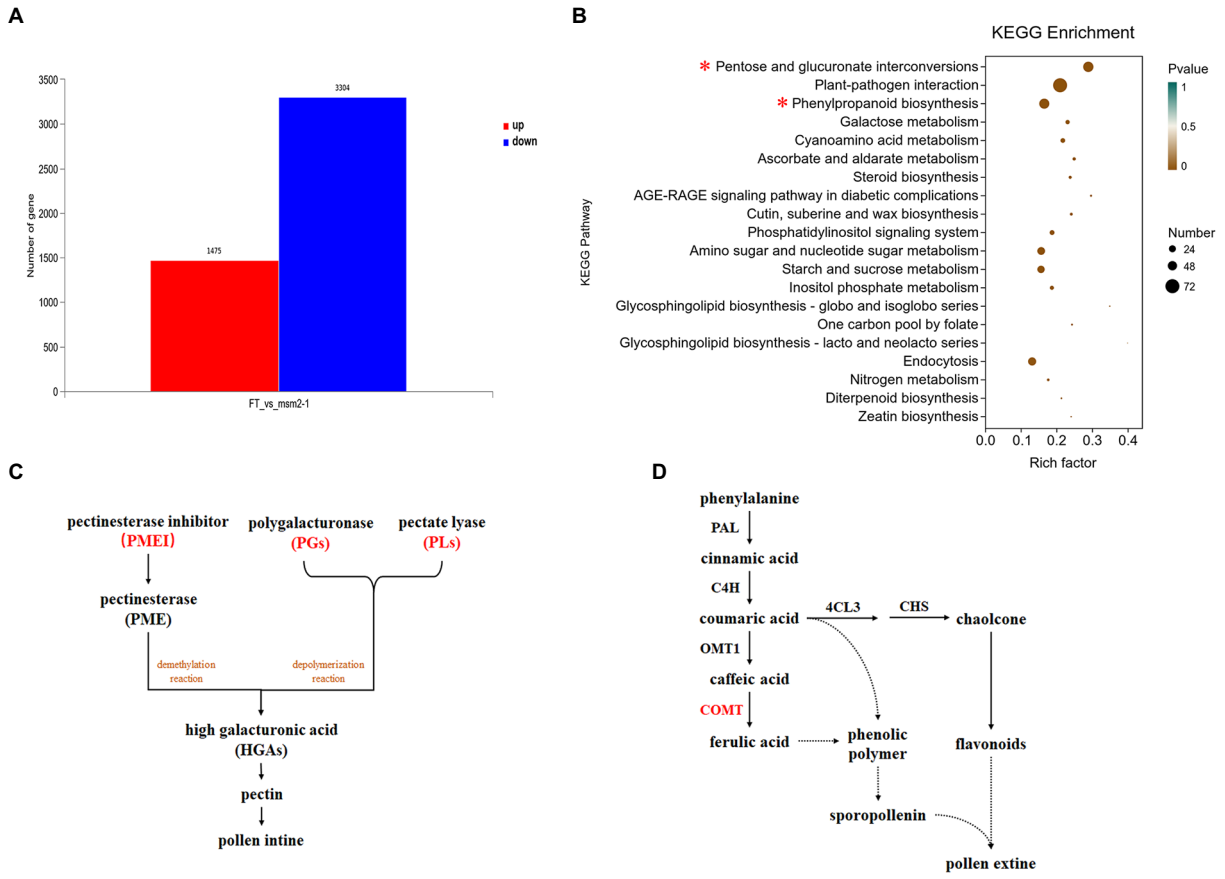


FIGURE 10
RNA-seq analysis between the wild-type "FT" and *msm2-1* mutant. **(A)** The number of different expression genes (DEGs) in the anthers between the wild-type "FT" and *msm2-1* mutant. **(B)** Kyoto Encyclopedia of Genes and Genomes enrichment results of the DEGs. **(C)** Schematic of pentose and glucuronate interconversions pathway related to pollen development. The red part represents the key enzymes enriched by the differentially expressed genes. **(D)** Schematic of phenylpropanoid metabolism pathway related to pollen development. The red part represents the key enzymes enriched by the differentially expressed genes.

development led to male sterile are diverse (Sanders et al., 1999; Zhang et al., 2011; Zhu et al., 2011; Zhang and Yang, 2014). Premature tapetum cell death caused the male sterile in rice *osmyb80* and *bm1* mutants (Pan et al., 2020; Xiang et al., 2021). Loss of *RGAT1* function also induced the premature degeneration of tapetal cells and pollen abortion in *Arabidopsis* (Qian et al., 2021). In addition, delayed tapetum degradation brought out defective mature pollen in rice *osmyb103* and *sts1* mutants (Lei et al., 2022; Yuan et al., 2022). In the present study, delayed tapetum degeneration is a hallmark defect in *msm2-1/2/3*, but the tapetum cells enlargement was not observed in *msm2-1/2/3* (Figures 2M–X). This phenotype was distinct from other tapetal PCD delayed mutants that have been described, such as *gamyb* and *tdr* mutants, which exhibited tapetal defects involving tapetum cell enlargement (Kaneko et al., 2004; Li et al., 2006), These findings suggested that the *BrMS1* mutation in *msm-1/2/3* mutants induced the delayed degradation of tapetum, but did not affect the morphology of tapetum cells.

Known tapetal PCD mutants also showed different types of pollen exine patterning. The manifestations of pollen exine abnormal in *msm2-1* mutant were different from other *MS1* homologous mutants. In the *Arabidopsis ms1* mutant, exine formation is abnormal, with initial primexine deposition but with only limited sporopollenin deposition (Ito and Shinozaki, 2002; Ariizumi et al., 2003; Vizcay-Barrena and Wilson, 2006). In the *rice*, the *osms1/ptc1* exines exhibited a two-layer structure, the sexine was abnormal because of the missing or decrease bacula (Yang et al., 2019b). In the *maize*, ubisch bodies were not enlarged in mutant *ms7-6007* and exine was much thinner than WT (An et al., 2020). In this study, the *msm2-1* mutant could form a complete pollen exine structure in the early stage of pollen development yet not observed the *PCPs* filled the pollen exine at the late stage. The excessive deposition of sporopollenin resulted in the pollen exine structure continuously thickened and formed a doughnut-like structure in the late stage of pollen development (Figures 4B,C,E,F). The SEM observation of pollen also exhibited a lack of elaborate patterning, giving rise to a smooth surface in

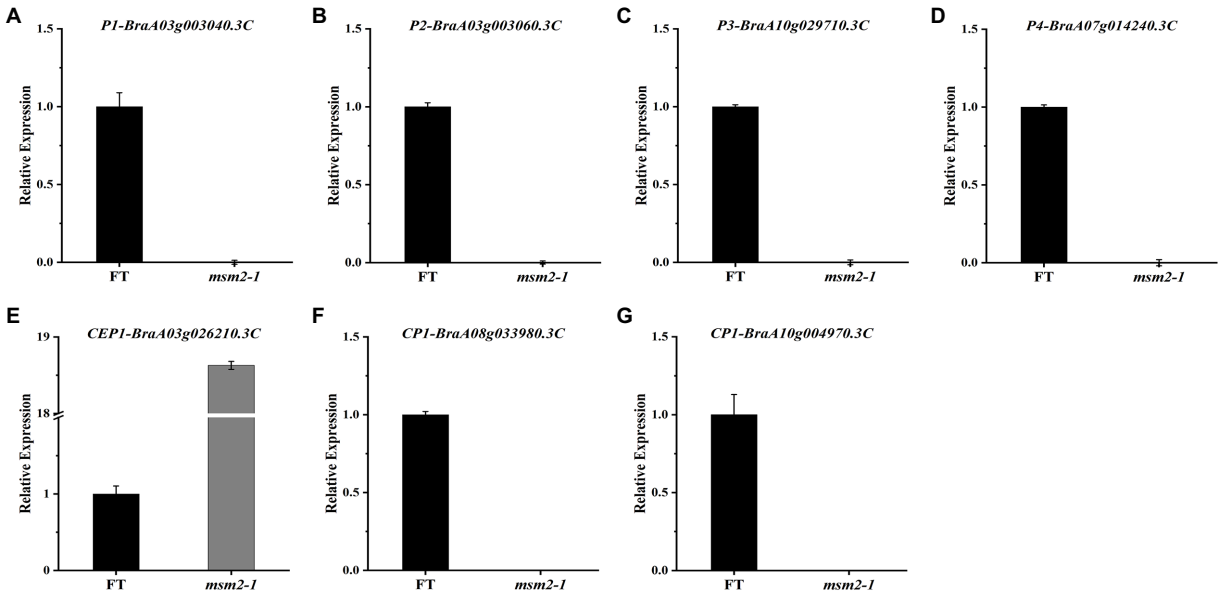


FIGURE 11
The quantitative reverse transcription polymerase chain reaction analysis of DEGs between the wild-type "FT" and *msm2-1* mutant. **(A–D)** The expression level of *pollen coat proteins* in anthers between the wild-type "FT" and *msm2-1* mutant. **(E–G)** The expression level of papain-like cysteine proteins in anthers between the wild-type "FT" and *msm2-1* mutant.

msm2-1 mutant (Figures 3C,D,G,H). In addition, abnormal morphology of pollen inner wall was observed in *msm2-1*. Taken together, loss of function of *BrMS1* results in delayed tapetal degeneration and pollen wall abnormal formation, and these developmental defects ultimately cause complete male sterility in the Chinese cabbage.

BrMS1 as a PHD-finger transcriptional activator negatively regulates its own expression

To decipher the molecular mechanism of the causal gene, MutMap and KASP analyses revealed that *BrMS1* (*BraA10g019050.3C*) encoded a PHD-finger transcriptional factor, was the causal gene in mutant *msm2-1* and confirmed function by *BrMS1* cloning in the allelic mutants *msm2-2/3* (Figures 5, 6). A previous study has shown that the PHD-finger motif was required for *MS1* function in *Arabidopsis* (Ito et al., 2007). The mutation of lacking the PHD-finger motif in both *OsMS1* and *PTC1*, caused male sterility in rice (Ito et al., 2007; Li et al., 2011; Yang et al., 2019b). Here, the base variations from G to A at 2443th of *msm2-1*, 1373th of *msm2-2* and 1887th of *msm2-3* made the PHD-finger domain abnormal in the allelic mutants and eventually led to pollen abortion (Figures 6A,B). Therefore, this demonstrates to some extent that the PHD-finger motif is indispensable for the development of anthers in Chinese cabbage.

Previous studies have implied *MS1/PTC1* could down-regulate their own expression in wild-type plants, and conjectured that *MS1* down-regulated its expression may due to conformational problem with the truncated mutant polypeptides and/or that the PHD-finger motif binding to its own promoter (Yang et al., 2007; Li et al., 2011). In this paper, qRT-PCR analysis revealed that *BrMS1* transcripts were increased to varying degrees in the *msm2-1/2/3* mutant, implying that self-regulatory based upon a normal gene functional could limit the expression of *BrMS1* in Chinese cabbage (Figure 8A). The expression of *BrMS1* in *msm2-1/2* allele mutants showed a greater expression compared with in *msm2-3*, we speculated that this was on account of the smaller difference in the 3D structure of the *BrMS1* gene in *msm2-3* versus the wild type "FT" (Figures 6C–F).

BrMS1 plays a key role in Chinese cabbage pollen development

To gain a better understanding of how *BrMS1* affects pollen development, we produced transcriptomes of the wild-type "FT" and the *msm2-1* mutant to identify potential genes and pathways involved in pollen development. The pollen coat rich in lipids and proteins plays an important role in the attachment and recognition of the pollen to the stigma (Kobayashi et al., 2021). The pollen coat is necessary for fertility, as pollen coat mutants in rice lead to the humidity-sensitive male sterility (Xue et al., 2018; Yu et al., 2019; Chen et al., 2020). In this study, RNA-seq profiling demonstrated that pollen coat

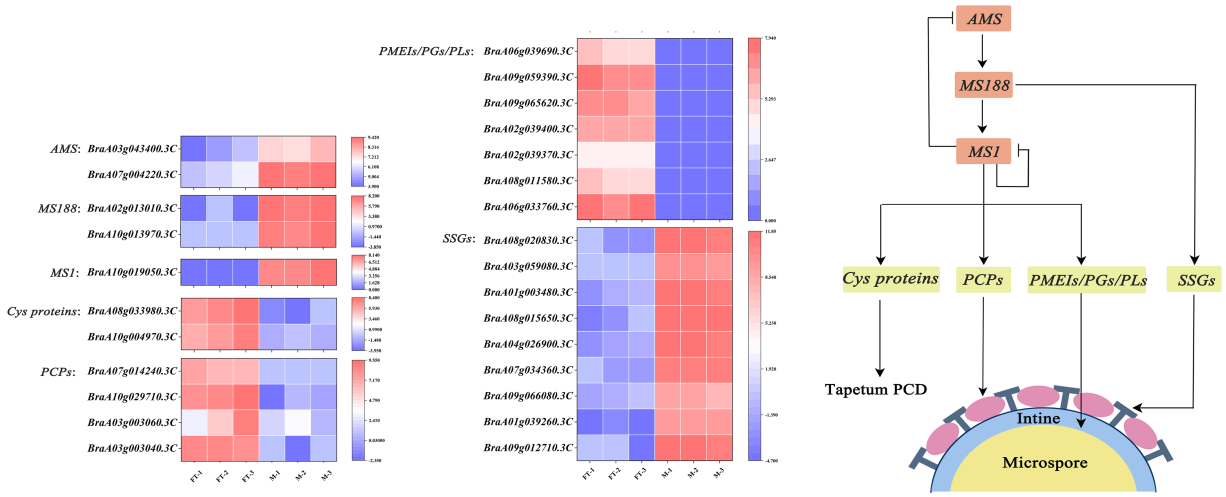


FIGURE 12
Proposed working model for *BrMS1* regulation of male fertility in Chinese cabbage. *MS1* which encodes a PHD-finger transcription factor could regulate the tapetum programmed cell death by influencing the transcription level of *Cys* proteins. *MS1* functions in pollen coats and pollen intine formation by mediating the expression of related genes in anthers. In addition, *MS1* could negatively regulate the transcription level of itself and the upstream regulator *AMS*, which in turn affects the expression of *SSGs* to control the formation of pollen exine. *MS1* plays an important role as a downstream regulator in the male fertility regulatory network.

proteins *GRPs* and *EXLs* significantly down-regulated in *msm2-1* compared with the wild-type "FT" (Supplementary Table S11). Accumulating reports implicated that papain-like cysteine proteases play a role in tapetal PCD, our results indicated that cysteine proteases *BrCEP1* (*BraA03g026210.3C*) was up-regulated and the other two cysteine proteases genes *BrCP1* (*BraA10g004970.3C* and *BraA08g033980.3C*) were down-regulated in *msm2-1* (Supplementary Table S13). These findings, to some extent, revealed a potential cause of the delayed degradation of tapetum in the *msm2-1* mutant. KEGG pathway analysis indicated that differentially expressed genes were significantly enriched in the pentose and glucuronate interconversions pathway. In this pathway, most DEGs involved in pectin biosynthesis, the main component of the pollen intine. *HGAs* are the most abundant polymer of pectin (Wolf et al., 2009). The HGA-modifying process can be strictly regulated by multiple cell wall enzymes. It has been demonstrated that the demethylesterification of *HGAs* could be spatiotemporally controlled by pectin methylesterases (*PMEs*) and PMEs activity might be regulated by a specific proteinaceous *PME* inhibitors (*PMEIs*; Camardella et al., 2000). Polygalacturonases (*PGs*) and pectin lyases (*PLs*) can depolymerize *HGAs* (Wakabayashi et al., 2003). Most of *PMEIs*, *PGs* and *PLs* were down-regulated in *msm2-1* (Figure 10C; Supplementary Table S9). Thus, we speculated that the defection of *BrMS1* regulated the development of pollen intine by affecting the expression of genes relevant to pectin biosynthesis in *msm2-1* mutant. However, the specific way in which *BrMS1* regulates these fertility-related genes will require further investigation.

Sporopollenin is the main component of the pollen exine and that its biosynthesis is closely related to fatty acid and phenylpropanoid metabolism (Liu and Fan, 2013; Shi et al., 2015). In this study, DEGs was likewise enriched in the phenylpropanoid metabolism pathway. The expression of some key enzymes in this pathway were significantly changed in *msm2-1* (Figure 10D; Supplementary Table S10). In addition, several *SSGs* were up-regulated expression in mutant *msm2-1*. This finding was consistent with the excessive deposition of sporopollenin of the *msm2-1* genotype (Figure 4; Supplementary Table S12). As we know that *MS188* directly regulates sporopollenin biosynthesis genes, and *MS188* is an upstream regulator of *MS1* in the tapetum regulatory network (*DYT1-TDF1-AMS-MS188-MS1*; Xiong et al., 2016; Wang et al., 2018). Herein, the expression level of *MS188* was up-regulated (Supplementary Table S13). In addition, several reports have shown that *MS1* could negatively regulate *AMS* expression, possibly through chromatin remodeling, or indirect protein degradation (Ferguson et al., 2017). This complex network of feedback loops is critical for the anther development. We suspected that mutation in *BrMS1* broke the balance of this negative feedback mechanism, resulting in the up-regulation expression of *AMS* and its downstream target gene *MS188*. While the *SSGs* directly regulated by the *MS188* were also up-regulated, giving rise to pollen exine abnormal in the mutant *msm2-1*. *BrMS1* as a TF could not only regulate the expression of downstream potential pollen development-related genes, but also effect on the expression of upstream genes (Figure 12). Hence, *BrMS1* plays an important role in

maintaining the stable and timely expression of various TFs in the tapetum regulatory network.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found at: https://www.ncbi.nlm.nih.gov/sra//PRJNA856064.

Author contributions

Data analysis and writing of the manuscript were performed by SD. Study conception and design were performed by SH and HF. JZ, YZ, FS, and GS carried out the experiments. BF took charge of English language editing. All authors contributed to the article and approved the submitted version.

Funding

The research was supported by the National Natural Science Foundation of China (grant no. 31730082).

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.2022.992391/full#supplementary-material

References

An, X. L., Ba, B., Duan, M. J., Dong, Z. Y., Liu, R. G., Yuan, D. Y., et al. (2020). Molecular regulation of ZmMs7 required for maize male fertility and development of a dominant male-sterility system in multiple species. *Proc. Natl. Acad. Sci. U. S. A.* 117, 23499–23509. doi: 10.1073/pnas.2010255117

Ariizumi, T., Hatakeyama, K., Hinata, K., Sato, S., Kato, T., Tabata, S., et al. (2003). A novel male-sterile mutant of Arabidopsis thaliana, faceless pollen-1, produces pollen with a smooth surface and an acetolysis-sensitive exine. *Plant Mol. Biol.* 53, 107–116. doi: 10.1023/B:PLAN.0000009269.97773.70

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Camardella, L., Carratore, V., Ciardiello, M., Servillo, L., Balestrieri, C., and Giovane, A. (2000). Kiwi protein inhibitor of pectin methylesterase amino-acid sequence and structural importance of two disulfide bridges. *Eur. J. Biochem.* 267, 4561–4565. doi: 10.1046/j.1432-1327.2000.01510.x

Chen, H., Zhang, Z., Ni, E., Lin, J., Peng, G., Huang, J., et al. (2020). HMS1 interacts with HMS1I to regulate very-long-chain fatty acid biosynthesis and the humidity-sensitive genic male sterility in rice (*Oryza sativa*). *New Phytol.* 225, 2077–2093. doi: 10.1111/nph.16288

Fan, Y. R., and Zhang, Q. F. (2018). Genetic and molecular characterization of photoperiod and thermo-sensitive male sterility in rice. *Plant Reprod.* 31, 3–14. doi: 10.1007/s00497-017-0310-5

Ferguson, A. C., Pearce, S., Band, L. R., Yang, C. Y., Ferjentsikova, I., King, J., et al. (2017). Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in *Arabidopsis*. *New Phytol.* 213, 778–790. doi: 10.1111/nph.14200

Gao, Y., Huang, S. N., Qu, G. Y., Fu, W., Zhang, M. D., Liu, Z. Y., et al. (2020). The mutation of ent-kaurene synthase, a key enzyme involved in gibberellin biosynthesis, confers a non-heading phenotype to Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Hortic Res.* 7, 178. doi: 10.1038/s41438-020-00399-6

Gao, Y., Qu, G., Huang, S., Liu, Z., Zhang, M., Fu, W., et al. (2022). Comparison between germinated seed and isolated microspore EMS mutagenesis in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Horticulturae* 8, 232. doi: 10.3390/horticulturae8030232

Hafidh, S., Fila, J., and Honys, D. (2016). Male gametophyte development and function in angiosperms: a general concept. *Plant Reprod.* 29, 31–51. doi: 10.1007/s00497-015-0272-4

Huang, S., Liu, Z., Li, C., Yao, R., Li, D., Hou, L., et al. (2017). Transcriptome analysis of a female-sterile mutant (fsm) in Chinese cabbage (*Brassica campestris* ssp. *pekinensis*). *Front Plant Sci.* 8, 546. doi: 10.3389/fpls.2017.00546

Huang, S., Peng, S., Liu, Z., Li, C., Tan, C., Yao, R., et al. (2020). Investigation of the genes associated with a male sterility mutant (msm) in Chinese cabbage (*Brassica campestris* ssp. *pekinensis*) using RNA-Seq. *Mol. Gen. Genomics.* 295, 233–249. doi: 10.1007/s00438-019-01618-z

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. (2007). *Arabidopsis* MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell* 19, 3549–3562. doi: 10.1105/tpc.107.054536

Ito, T., and Shinozaki, K. (2002). The MALE STERILITY1 gene of *Arabidopsis*, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. *Plant Cell Physiol.* 43, 1285–1292. doi: 10.1093/pcp/pcf154

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Kaneko, M., Inukai, Y., Ueguchi-Tanaka, M., Itoh, H., Izawa, T., Kobayashi, Y., et al. (2004). Loss-of-function mutations of the rice GAMYB gene impair α-amylase expression in aleurone and flower development. *Plant Cell* 16, 33–44. doi: 10.1105/tpc.017327

Kim, Y. J., and Zhang, D. B. (2018). Molecular control of male fertility for crop hybrid breeding. *Trends Plant Sci.* 23, 53–65. doi: 10.1016/j.tplants.2017.10.001

Kobayashi, K., Akita, K., Suzuki, M., Ohta, D., and Nagata, N. J. P. B. (2021). Fertile *Arabidopsis* cyp 704b1 mutant, defective in sporopollenin biosynthesis, has a normal pollen coat and lipidic organelles in the tapetum. *Plant Biotechnol.* 38, 109–116. doi: 10.5511/plantbiotechnology.20.1214b

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lei, T., Zhang, L. S., Feng, P., Liu, Y., Yin, W. Z., Shang, L. N., et al. (2022). OsMYB103 is essential for tapetum degradation in rice. *Theor. Appl. Genet.* 135, 929–945. doi: 10.1007/s00122-021-04007-6

Li, H., Yuan, Z., Vizcay-Barrena, G., Yang, C. Y., Liang, W. Q., Zong, J., et al. (2011). PERSISTENT TAPETAL CELL1 encodes a PHD-finger protein That is

required for Tapetal CELL death and pollen development in Rice. *Plant Physiol.* 156, 615–630. doi: 10.1104/pp.111.175760

Li, N., Zhang, D.-S., Liu, H.-S., Yin, C.-S., Li, X. X., Liang, W. Q., et al. (2006). The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell* 18, 2999–3014. doi: 10.1105/tpc.106.044107

Liang, J., Ma, Y., Wu, J., Cheng, F., Liu, B., and Wang, X. (2017). Map-based cloning of the dominant genic male sterile Ms-cd1 gene in cabbage (Brassica oleracea). *Theor. Appl. Genet.* 130, 71–79. doi: 10.1007/s00122-016-2792-1

Lin, S., Dong, H., Zhang, F., Qiu, L., Wang, F. Z., Cao, J. S., et al. (2014). BcMF8, a putative arabinogalactan protein-encoding gene, contributes to pollen wall development, aperture formation and pollen tube growth in *Brassica campestris*. *Ann. Bot.* 113, 777–788. doi: 10.1093/aob/mct315

Lin, T.-K., Lin, Y.-P., and Lin, S.-F. (2020). Genetic analysis and fine mapping of a spontaneously mutated male sterility gene in *Brassica rapa* ssp. *chinensis*. *G3* 10, 1309–1318. doi: 10.1534/g3.120.401091

Liu, L., and Fan, X. D. (2013). Tapetum: regulation and role in sporopollenin biosynthesis in *Arabidopsis*. *Plant Mol. Biol.* 83, 165–175. doi: 10.1007/s11103-013-0085-5

Liu, H., Hu, M., Wang, Q., Cheng, L., and Zhang, Z. (2018). Role of papain-like cysteine proteases in plant development. *Front. Plant Sci.* 9:1717. doi: 10.3389/fpls.2018.01717

Liu, Q., Lan, Y., Wen, C., Zhao, H., Wang, J., and Wang, Y. (2016). Transcriptome sequencing analyses between the cytoplasmic male sterile line and its maintainer line in welsh onion (*Allium fistulosum* L.). *Int. J. Mol. Sci.* 17:1058. doi: 10.3390/ijms17071058

Lu, J. Y., Xiong, S. X., Yin, W. Z., Teng, X. D., Lou, Y., Zhu, J., et al. (2020). MS1, a direct target of MS188, regulates the expression of key sporophytic pollen coat protein genes in *Arabidopsis*. *J. Exp. Bot.* 71, 4877–4889. doi: 10.1093/jxb/eraa219

Ma, C. F., Zhu, C. Z., Zheng, M., Liu, M. C., Zhang, D. J., Liu, B. L., et al. (2019). CRISPR/Cas 9-mediated multiple gene editing in Brassica oleracea var. capitata using the endogenous tRNA-processing system. *Hortic Res.* 6, 20. doi: 10.1038/s41438-018-0107-1

Pan, X., Yan, W., Chang, Z., Xu, Y., Luo, M., Xu, C., et al. (2020). OsMYB80 regulates anther development and pollen fertility by targeting multiple biological pathways. *Plant Cell Physiol.* 61, 988–1004. doi: 10.1093/pcp/pcaa025

Parish, R. W., and Li, S. F. (2010). Death of a tapetum: a programme of developmental altruism. *Plant Sci.* 178, 73–89. doi: 10.1016/j.plantsci.2009.11.001

Preston, J., Wheeler, J., Heazlewood, J., Li, S. F., and Parish, R. W. J. T. P. J. (2004). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Plant J.* 40, 979–995. doi: 10.1111/j.1365-313X.2004.02280.x

Qian, Q., Yang, Y., Zhang, W., Hu, Y., Li, Y., Yu, H., et al. (2021). A novel *Arabidopsis* gene RGAT1 is required for GA-mediated tapetum and pollen development. *New Phytol.* 231, 137–151. doi: 10.1111/nph.17314

Sanders, P. M., Bui, A. Q., Weterings, K., McIntire, K. N., Hsu, Y.-C., Lee, P. Y., et al. (1999). Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sex. Plant Reprod.* 11, 297–322. doi: 10.1007/s004970050158

Sharma, P., Kaur, M., Sharma, A., and Bhardwaj, N. (2021). Breeding vegetables for protected cultivation: A review. *Himachal J. Agric. Res.* 47, 1–17.

Shi, J., Cui, M., Yang, L., Kim, Y. J., and Zhang, D. (2015). Genetic and biochemical mechanisms of Pollen Wall development. *Trends Plant Sci.* 20, 741–753. doi: 10.1016/j.tplants.2015.07.010

Singh, S., Dey, S. S., Bhatia, R., Kumar, R., and Behera, T. K. (2019). Current understanding of male sterility systems in vegetable brassicas and their exploitation in hybrid breeding. *Plant Reprod.* 32, 231–256. doi: 10.1007/s00497-019-00371-y

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

Tan, C., Liu, Z., Huang, S., and Feng, H. (2019). Mapping of the male sterile mutant gene ftms in *Brassica rapa* L. ssp. *pekinensis* via BSR-Seq combined with whole-genome resequencing. *Theor. Appl. Genet.* 132, 355–370. doi: 10.1007/s00122-018-3223-2

Vizcay-Barrena, G., and Wilson, Z. (2006). Altered tapetal PCD and pollen wall development in the *Arabidopsis* ms1 mutant. *J. Exp. Bot.* 57, 2709–2717. doi: 10.1093/jxb/erl032

Wakabayashi, K., Hoson, T., and Huber, D. J. (2003). Methyl de-esterification as a major factor regulating the extent of pectin depolymerization during fruit ripening: a comparison of the action of avocado (Persea americana) and tomato (Lycopersicon esculentum) polygalacturonases. *J. Plant Physiol.* 160, 667–673. doi: 10.1078/0176-1617-00951

Wan, X. Y., Wu, S. W., Li, Z. W., An, X. L., and Tian, Y. H. (2020). Lipid metabolism: critical roles in male fertility and other aspects of reproductive development in plants. *Mol. Plant* 13, 955–983. doi: 10.1016/j.molp.2020.05.009

Wan, X. Y., Wu, S. W., Li, Z. W., Dong, Z. Y., An, X. L., Ma, B., et al. (2019). Maize genic male-sterility genes and their applications in hybrid breeding: progress and perspectives. *Mol. Plant* 12, 321–342. doi: 10.1016/j.molp.2019.01.014

Wang, K., Guo, Z. L., Zhou, W. T., Zhang, C., Zhang, Z. Y., Lou, Y., et al. (2018). The regulation of sporopollenin biosynthesis genes for rapid Pollen Wall formation. *Plant Physiol.* 178, 283–294. doi: 10.1104/pp.18.00219

Wang, Y., Ye, H., Bai, J., and Ren, F. (2021). The regulatory framework of developmentally programmed cell death in floral organs: a review. *Plant Physiol. Biochem.* 158, 103–112. doi: 10.1016/j.plaphy.2020.11.052

Wang, N., Zhang, Y., Huang, S., Liu, Z., Li, C., and Feng, H. (2020). Defect in Brnym1, a magnesium-dechelatase protein, causes a stay-green phenotype in an EMS-mutagenized Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) line. *Hortic Res.* 7, 8. doi: 10.1038/s41438-019-0223-6

Wolf, S., Mouille, G., and Pelloux, J. (2009). Homogalacturonan methyl-esterification and plant development. *Mol. Plant* 2, 851–860. doi: 10.1093/mp/ssp066

Wu, H.-M., and Cheung, A. Y. (2000). Programmed cell death in plant reproduction. *Plant Mol. Biol.* 44, 267–281. doi: 10.1023/A:1026536324081

Xi, X., Wei, K. Y., Gao, B. Z., Liu, J. H., Liang, J. L., Cheng, F., et al. (2018). BrFLC5: a weak regulator of flowering time in *Brassica rapa*. *Theor. Appl. Genet.* 131, 2107–2116. doi: 10.1007/s00122-018-3139-x

Xiang, X., Sun, L., Yu, P., Yang, Z., Zhang, P., Zhang, Y., et al. (2021). The MYB transcription factor Baymax1 plays a critical role in rice male fertility. *Theor. Appl. Genet.* 134, 453–471. doi: 10.1007/s00122-020-03706-w

Xiong, S. X., Lu, J. Y., Lou, Y., Teng, X. D., Gu, J. N., Zhang, C., et al. (2016). The transcription factors MS 188 and AMS form a complex to activate the expression of CYP 703A2 for sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 88, 936–946. doi: 10.1111/tpj.13284

Xue, Z., Xu, X., Zhou, Y., Wang, X., Zhang, Y., Liu, D., et al. (2018). Deficiency of a triterpene pathway results in humidity-sensitive genic male sterility in rice. *Nat. Commun.* 9, 1–10. doi: 10.1038/s41467-018-03048-8

Yang, Z. F., Liu, L., Sun, L. P., Yu, P., Zhang, P. P., Abbas, A., et al. (2019a). OsMS1 functions as a transcriptional activator to regulate programmed tapetum development and pollen exine formation in rice. *Plant Mol. Biol.* 99, 175–191. doi: 10.1007/s11103-018-0811-0

Yang, Z., Sun, L., Zhang, P., Zhang, Y., Yu, P., Liu, L., et al. (2019b). TDR INTERACTING PROTEIN 3, encoding a PHD-finger transcription factor, regulates Ubisch bodies and pollen wall formation in rice. *Plant J.* 99, 844–861. doi: 10.1111/tpj.14365

Yang, C., Vizcay-Barrena, G., Conner, K., and Wilson, Z. A. (2007). MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell* 19, 3530–3548. doi: 10.1105/tpc.107.054981

Yoo, S., Cho, Y., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199

Yu, B., Liu, L., and Wang, T. J. P. (2019). Deficiency of very long chain alkanes biosynthesis causes humidity-sensitive male sterility via affecting pollen adhesion and hydration in rice. *Plant Cell Environ.* 42, 3340–3354. doi: 10.1111/pce.13637

Yuan, G., Zou, T., He, Z., Xiao, Q., Li, G., Liu, S., et al. (2022). SWOLLEN TAPETUM AND STERILITY 1 is required for tapetum degeneration and pollen wall formation in rice. *Plant Physiol.* kiac307. doi: 10.1093/plphys/kiac307 [Epub ahead of print].

Zhang, X. J., Chen, H. Y., Zhang, Q., Zhang, Y. X., Xiao, Z. D., Guo, Y., et al. (2020). Cytological and genetic characterisation of dominant GMS line Shaan-GMS inBrassica napusL. *J. Appl. Genet.* 61, 477–488. doi: 10.1007/s13353-020-00570-8

Zhang, D., Liu, D., Lv, X., Wang, Y., Xun, Z., Liu, Z., et al. (2014). The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in *Arabidopsis*. *Plant Cell* 26, 2939–2961. doi: 10.1105/tpc.114.127282

Zhang, D. B., Luo, X., and Zhu, L. (2011). Cytological analysis and genetic control of rice anther development. *J. Genet. Genomics* 38, 379–390. doi: 10.1016/j.jgg.2011.08.001

Zhang, D., Shi, J., and Yang, X. (2016). "Role of lipid metabolism in plant pollen Exine development," in *Lipids in Plant and Algae Development*. eds. Y. Nakamura and Y. Li-Beisson (Cham: Springer International Publishing), 315–337.

Zhang, D., and Yang, L. (2014). Specification of tapetum and microsporocyte cells within the anther. *Curr. Opin. Plant Biol.* 17, 49–55. doi: 10.1016/j.pbi.2013.11.001

Zhang, Z. B., Zhan, H. D., Lu, J. Y., Xiong, S. X., Yang, N. Y., Yuan, H. Y., et al. (2021). Tapetal 3-ketoacyl-coenzyme a synthases are involved in pollen coat lipid

accumulation for pollen-stigma interaction in *Arabidopsis*. *Front. Plant Sci.* 12:770311. doi: 10.3389/fpls.2021.770311

Zhang, Z., Zhu, J., Gao, J., Wang, C., Li, H., Li, H., et al. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *Plant J.* 52, 528–538. doi: 10.1111/j.1365-313X.2007.03254.x

Zhou, X., Liu, Z. Y., Ji, R. Q., and Feng, H. (2017). Comparative transcript profiling of fertile and sterile flower buds from multiple-allele-inherited male sterility in Chinese cabbage (*Brassica campestris* L. ssp *pekinensis*). *Mol. Gen. Genomics.* 292, 967–990. doi: 10.1007/s00438-017-1324-2

Zhu, J., Chen, H., Li, H., Gao, J.-F., Jiang, H., Wang, C., et al. (2008). Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *Plant J.* 55, 266–277. doi: 10.1111/j.1365-313X.2008.03500.x

Zhu, J., Lou, Y., Xu, X. F., and Yang, Z. N. (2011). A genetic pathway for Tapetum development and function in *Arabidopsis*. *J. Integr. Plant Biol.* 53, 892–900. doi: 10.1111/j.1744-7909.2011.01078.x

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpls.2022.898108

Fine Mapping of Clubroot Resistance Loci *CRA8.1* and Candidate Gene Analysis in Chinese Cabbage (*Brassica rapa* L.)

Yanyan Wang[1], Xianyu Xiang[1], Fan Huang[1], Wenlin Yu[1], Xueqing Zhou[1], Baojun Li[1,2], Yunyun Zhang[1,3], Peng Chen[1] and Chunyu Zhang[1]**

[1] College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2] Hybrid Rape Research Center of Shaanxi Province, Shaanxi Rapeseed Branch of National Centre for Oil Crops Genetic Improvement, Yangling, China, [3] Industrial Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China

Clubroot is caused by *Plasmodiophora brassicae*, which threatens *Brassicaceae* crop production worldwide. In recent years, there has been an outbreak and rapid spread of clubroot in many major cruciferous crop-producing areas of China. In this study, we identified a cabbage material DingWen (DW) with different resistant capabilities from Huashuang5R (H5R) and Huayouza62R of *Brassica napus*, which are currently used as the main resistant cultivars for clubroot management in China. We used a next-generation sequencing-based bulked segregant analysis approach, combined with genetic mapping to identify clubroot-resistant (CR) genes from F_1 population generated from a cross between the DW (CR) and HZSX (clubroot susceptible). The CR locus of DW (named *CRA8.1*) was mapped to a region between markers A08-4346 and A08-4853, which contains two different loci *CRA8.1a* and *CRA8.1b* after fine mapping. The *CRA8.1b* loci contain a fragment of 395 kb between markers A08-4624 and A08-4853 on A08 chromosome, and it is responsible for the resistance to *PbZj* and *PbXm* isolates. However, together with *CRA8.1a,* corresponding to a 765-kb region between markers A08-4346 and A08-4624, then it can confer resistance to *PbXm*$^{+}$. Finally, through expression analysis between resistant and susceptible materials, two genes encoding TIR-NBS-LRR proteins (*BraA08g039211E* and *BraA08g039212E*) and one gene encoding an RLP protein (*BraA08g039193E*) were identified to be the most likely CR candidates for the peculiar resistance in DW.

Keywords: clubroot, *Plasmodiophora brassicae*, *Brassica rapa*, *CRA8.1*, fine mapping

OPEN ACCESS

Edited by:
Kun Lu,
Southwest University, China

Reviewed by:
Ying Fu,
Zhejiang Academy of Agricultural Sciences, China
Wei Qian,
Southwest University, China

***Correspondence:**
Peng Chen
chenpeng@mail.hzau.edu.cn
Chunyu Zhang
zhchy@mail.hzau.edu.cn

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

Received: *17 March 2022*
Accepted: *07 April 2022*
Published: *06 May 2022*

Citation:
Wang Y, Xiang X, Huang F, Yu W, Zhou X, Li B, Zhang Y, Chen P and Zhang C (2022) Fine Mapping of Clubroot Resistance Loci CRA8.1 and Candidate Gene Analysis in Chinese Cabbage (Brassica rapa L.). Front. Plant Sci. 13:898108. doi: 10.3389/fpls.2022.898108

INTRODUCTION

Clubroot is a highly contagious soil-borne disease that specifically endangers *Brassica* species, and it is caused by *Plasmodiophora brassicae*, a pathogen that belongs to peculiar eukaryotic taxa under Endomyxa branch (Dixon, 2009). Most of the Brassica plants, including *Brassica napus*, *Brassica oleracea*, and *Brassica rapa*, can be infected by *P. brassicae*. The susceptible plants display a swollen root symptom by this disease infection, which leads to dysfunction of root function for water and nutrient transport, eventually withering and death (Gahatraj et al., 2019). In recent years, the

disease is rapidly spreading from different continents and countries, resulting in severe yield loss of cruciferous crops and vegetables (Hwang et al., 2012).

Although biological and chemical reagents can rescue the disease phenotype under certain conditions, the identification and introgression of resistance genes to create resistant varieties is the most efficient and economic approach to defeat the disease spreading (Elke et al., 2009; Chai et al., 2014). A few qualitative and quantitative clubroot-resistant (CR) loci have been identified in *B. rapa*, *B. oleracea*, *B. napus,* and *Brassica nigra* (Mehraj et al., 2020). The majority of CR genes are located on A genomes, including A01, A02, A03, A05, A06, and A08 chromosomes (Mehraj et al., 2020). In particular, CR loci on A03 or A08 chromosomes are extensively studied, including *PbBa3.1*, *PbBa3.2*, and *PbBa3.3*, *CRd*, *Crr3*, *CRk*, *CRb*, *Rcr4*, *Rcr1,* and *CRa* from A03 chromosome (Matsumoto et al., 1998; Hirai et al., 2004; Piao et al., 2004; Sakamoto et al., 2008; Chen et al., 2013; Chu et al., 2014; Yu et al., 2017; Pang et al., 2018), and *Crr1* (including both *Crr1a* and *Crr1b*), *PbBa8.1*, *Rcr3, Rcr9wa*, *BraA.CR.b,* and *Rcr9* from A08 chromosome (Suwabe et al., 2003; Chen et al., 2013; Yu et al., 2017; Hirani et al., 2018; Karim et al., 2020). *CrrA5* and *Crr4* were located on A05 and A06 chromosomes (Suwabe et al., 2006; Nguyen et al., 2018), and *Crr2, PbBa1.1*, *CRc*, and *Rcr8* were mapped to A01 and A02 chromosomes, respectively (Suwabe et al., 2003; Sakamoto et al., 2008; Chen et al., 2013; Yu et al., 2017). Based on fine-mapping markers, some CR loci reside extremely close to or overlap each other, such as *Crr1* and *PbBa8.1*, and *CRa* and *CRb*, and they can be allelic and actually originate from homologous genes from different A-genome species (Chen et al., 2013). These genes can be used as resources for breeding CR accessions.

Actually, from a broader view of disease resistance genes, most plant resistance (R) genes identified so far encode for cell membrane or intracellular receptors, including RLPs/RLKs (receptor-like proteins/kinases) and NLRs (nucleotide-binding site–leucine-rich repeat) (Kourelis and van der Hoorn, 2018). Plant innate immunity system comprises PTI (PAMP-triggered immunity) and ETI (effector-triggered immunity) (Deng et al., 2020). It is believed that specific interactions between a resistance gene and certain pathotypes are dependent mainly *via* ETI, but also affected by PTI pathway (Kim et al., 2016; Neik et al., 2017; Larkan et al., 2020; Yuan et al., 2021). For PTI, PAMPs from the invading pathogen are perceived by host membrane receptors known as PRRs (pattern recognition receptors), and RLKs and RLPs are the two major groups of PRRs on the cell surface determining the initial recognition (Wu and Zhou, 2013; Larkan et al., 2015; Grund et al., 2019; Kim et al., 2020). For the ETI route, effectors released inside the host cell are recognized by plant intracellular proteins such as NLRs, some of which require additional helper NLRs or chaperon proteins for triggering the ETI response (Grund et al., 2019). Different types of NLR proteins have been identified based on their protein domain structure, such as TIR-NBS-LRR, CC-NBS-LRR, and CC_R-NBS-LRR subgroups (Wang and Chai, 2020). The above-mentioned CR genes, only two CR genes, namely, *CRa* and *Crr1a*, have been successfully isolated, both encoding TIR-NBS-LRR (NLRs) proteins (Ueno et al., 2012; Hatakeyama et al., 2013).

Clubroot-resistant genes from *B. rapa* have been introgressed into susceptible *B. napus* in order to generate resistant varieties. Currently, the two CR *B. napus* varieties in China are Huashuang5R (H5R) and Hayouza62R, which were generated through crosses using resistant *B. rapa* donor parents ECD04 (AA, 2n = 20) and Shinki (AA, 2n = 20), respectively (Zhan et al., 2020; Li et al., 2021). They were resistant to most of the clubroot pathogens isolated from different regions in China (Nadil et al., 2019). However, due to the soil-borne nature of the clubroot pathogen, a "resistant" Brassica variety to one pathogen can become "susceptible" to another pathogen from a different source. The present clubroot pathogen classification systems include the "Williams" system (Williams, 1966), the ECD system (Buczacki et al., 1975), and the SCD system (Pang et al., 2020). According to the "Williams" system, pathotype 4 is the most prevalent form in China, and most of the isolates such as *PbZj* (Zhijiang isolated from Hubei province) and *PbXm* (Xinmin from Liaoning province) belong to this group. The two *B. napus* CR varieties H5R and Hayouza62R containing *PbBa8.1*- and *CRb*-resistant loci, respectively, are resistant to *PbZj* and *PbCd* (Chengdu) isolates, but not to *PbXm* pathogen (Shah et al., 2020). Another struggling problem is the loss of resistance over time, which also calls for a new resource of resistance genes. During previous work, people found that pyramiding multiple CR genes in one variety would greatly improve the resistance of the host plant. For example, *CRb* and *PbBa8.1* pyramiding lines in *B. napus* have demonstrated that plants in a homozygous state in each CR gene/locus had higher resistance than those in a heterozygous state (Shah et al., 2020). However, the fast spreading of clubroot disease, the presence of various pathotypes for the clubroot pathogens from different regions, and the frequent loss of resistance collectively call for urgent identification of a new CR gene resource.

The purpose of this study was to identify a CR locus different from *PbBa8.1* and *CRb* and to identify the candidate genes responsible for disease resistance, therefore providing foundation for future utilization of these new CR genes for resistance breeding in *Brassica* species.

MATERIALS AND METHODS

Plant Materials

The Brassica cultivars used in the current study include H5R, Huashuang5S (H5S), and 409R (the resistant restore line for Huayouza62R) of *B. napus*, DingWen (DW), HuangZiShaXun (HZSX, Z1) (Belser et al., 2018), and 91-12 of *B. rapa*. H5R and 409R are resistant cultivars containing CR loci of *PbBa8.1* and *CRb*, respectively. H5S and 91-12 were used as susceptible controls. DW is a hybrid and sterile resistant material. A susceptible cultivar HZSX was used as a parent in crosses with DW for F_1 population construction and gene mapping.

Plasmodiophora brassicae Collection and Clubroot Resistance Phenotyping

Different species of *P. brassicae* from the Chinese cabbage roots with severe galling were collected from five different locations

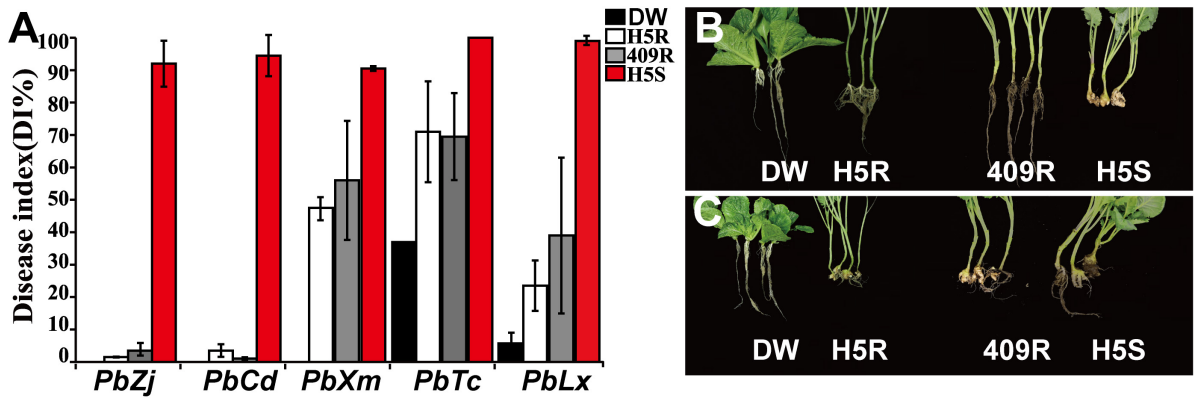


FIGURE 1 | Disease resistance responses of different materials to *P. brassicae* from different field isolates. **(A)** Disease index statistics. Black, white, gray, and red bars represent the disease index in DW, Huashuang5R (H5R), 409R, and Huashuang5S (H5S), respectively. **(B,C)** The phenotypes of DW, H5R, 409R, and H5S were inoculated with *PbZj* and *PbXm*, respectively.

of China: *PbXm*, Xinmin of Liaoning province; *PbCd*, Chengdu of Sichuan province; *PbZj*, Zhijang of Hubei province; *PbTc*, Tengchong of Yunnan province; and *PbLx*, Linxiang of Yunnan province. During the study, we found that Xinmin region contained two isolates, based on the host response. Therefore, we named it as "*PbXm*$^{+}$," in order to discriminate it from the other one. Galls were frozen and stored in −20°C until further use.

Clubroot resistance test of all materials was carried out as described previously (Chen et al., 2015). In brief, the collected galls were thawed at room temperature, ground to homogenization with buffer, and filtered through gauze, and resting spores were isolated after rounds of centrifugation. The spore concentration was measured and diluted to 10^7 resting spores per milliliter in sterile distilled water, before being used for host plant inoculation.

Host plants were prepared by germination of seeds on moist filter paper for 4 days, followed by the growth of seedling in a nursery room on medium at 16-h photoperiod, with 25/20°C day/night temperature for 7 days. When seedlings are ready, 1 ml of the above resting spore suspension was inoculated into the root of the seedling, and the disease symptoms were scored after 30–40 days. Clubroot severity was scored as 0–3 based on the root morphology, with grade 0 as normal growth, grade 1 as a few small galls on the lateral roots, grade 2 as big galls on lateral roots, and grade 3 as big galls on both primary roots and lateral roots. The disease index (DI) was calculated based on the number of plants at various severity levels (Shah et al., 2020).

Bulked Pool Construction and Sequencing

Bulked segregation analysis (BSA) method was used to map the genomic region responsible for CR in DW. A genetic cross was performed using DW and HZSX, and the two parents were genetically variant and also showed contrasting CR phenotype. Thirty-eight highly resistant and 38 extremely susceptible plants were selected from F_1 population to construct two pools (resistant pool and susceptible pool), respectively. Genomic DNA was extracted from leaves of individual plants using DNA Secure Plant Kit (TIANGEN, China). The DNA concentration was quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, United States), and DNA of each individual plant within the same pool is mixed in equal concentration, resulting in two DNA libraries for sequencing.

Second-generation sequencing was performed on these two libraries using an Illumina HiSeq platform in PE150 mode. The raw data were accessed for quality control and then transformed into clean data by removing adaptor sequence. Clean reads were mapped to the reference genome of *B. rapa*[1] using the Burrows–Wheeler Aligner (BWA0.7.12-r1039 mem).

CRA8.1 Gene Mapping

To identify candidate CR loci in DW, we first calculated the Euclidean distance (ED) value of SNP and InDel separately and then constructed maps based on the distribution of ED value on the chromosome. The molecular markers, developed by Zhan et al. (2020) on the A08 chromosome, were applied for genotyping. Five hundred F_1 plants were used in the primary mapping, and the population was expanded to 3290 F_1 plants for fine mapping. Physical map was constructed according to the location of molecular markers using MapChart software (Voorrips, 2002).

Assembly of *CRA8.1* Fragment

To get sequence information of fragment containing *CRA8.1* in DW, we performed fragment assembly with a European turnip ECD04 genome as the reference.

First, based on a Nova Seq 6000 sequencing platform, around 100 Gb Illumina short reads were obtained with leaf DNA of disease-resistant individual R59, which is the resistant plant in F_1. Then, the Illumina short reads were mapped to the reference genome by BWA software package and rearranged by Samtools software package with parameters all set as default

[1]http://39.100.233.196:82/download_genome/Brassica_Genome_data/Brara_Chiifu_V3.0/

TABLE 1 | Disease phenotype statistics in F_1 population from cross of DW and HZSX.

Parents and crosses	Type	Disease levels				Total	IR[a]	χ^2
		0	1	2	3			
DW	R-parent	30	0	0	0	30	0	
HZSX	S-parent	0	0	0	30	30	100	
DW × HZSX	F_1	69	0	0	70	139	50	0.007

[a]*Incidence rate (%).*

(Li and Durbin, 2009). Second, SNP/InDels data were extracted from the mapping results by BCF tools software package with the parameters set as "DP > 30 and QUAL > 100" (Narasimhan et al., 2016). At the same time, based on the sequence of A08-4346 and A08-5076 markers and blast result from NCBI[2] from the reference genome, chromosomal fragments between the two markers from the reference genome were extracted by bedtools software package with default parameters (Quinlan and Hall, 2010).

Then, based on SNP/InDels data, an R script was written to replace the nucleotide sequence of candidate fragment of DW. The non-heterozygous SNP/InDels loci were replaced directly, while for loci containing heterozygous SNP/InDels, only one of the SNP/InDels was randomly selected to replace the candidate sequence.

Gene Synteny Analysis of *CRA8.1* Fragment

First, we downloaded the Chinese cabbage HuangZiShaXun (Z1) genome,[3] *B. napus* ZS11 genome,[4] and European turnip ECD04 genome (unpublished). Then, the CDS of each genome was extracted by GFF read software package with parameter set as default. Finally, the gene synteny analysis was performed with CDS of candidate fragment as query and CDS of each genome as subject with MCscan.[5]

Identification and Expression Analysis of Candidate Genes

We used DW and HZSX (susceptible parent) and inoculated them with *PbZj* or *PbXm*$^+$ isolates, respectively. Root samples were taken at 0 h, 12 h, and 4 days after inoculation, total RNA was extracted, and the candidate gene expression was analyzed using qRT-PCR. Information of qRT-PCR primers is listed in **Supplementary Table 1**, and *BrUBC10* (XM_009134237.3) was used as an internal reference.

For promoter *cis*-element analysis, the DNA sequence 3.0 kb upstream of the coding sequence was submitted to the Plant CARE database[6] for the prediction of putative *cis*-elements. The data were presented using TBtools software.

[2]https://blast.ncbi.nlm.nih.gov/Blast.cgi

[3]http://www.genoscope.cns.fr/plants

[4]http://cbi.hzau.edu.cn/bnapus/

[5]https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)

[6]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

RESULTS

Chinese Cabbage "DW" With Broad Resistance to *Plasmodiophora brassicae* From Different Regions

Previously, a Chinese cabbage material was screened out with excellent resistance to Zj isolate of *P. brassicae*, and this material was named "DW." In order to further verify the pathogen resistance profile of DW and characterize the identity of CR genes in comparison with the ones currently widely used in H5R and Huayouza62R *B. napus* varieties in China, DW, H5R, 409R (resistant, restore line of Huayouza62R), and H5S were inoculated by five different clubroot isolates collected from different locations of China, namely, *PbXm*, *PbCd*, *PbZj*, *PbTc,* and *PbLx* (**Figure 1A**). Indoor inoculation experiments showed that susceptible control H5S was severely infected by all pathogens, while H5R, 409R, and DW had variable resistance (**Figure 1A**). H5R, 409R, and DW showed high resistance to *PbZj* and *PbCd* (DI $\leq$ 3.5%) (**Figures 1A,B**), and DW was the only one resistant to *PbXm*, whereas H5R and 409R showed a DI level of 47.5 and 56%, respectively (**Figures 1A,C**). We can also see that all four materials used in this study were susceptible to *PbTc*; however, DW displayed lower DI than H5R and 409R (**Figure 1A**). Similarly, H5R and 409R showed 24–40% DI to *PbLx*, but the DI of DW was below 10% (**Figure 1A**). Taken together, these results suggest that DW is a material with superior CR, which is fundamentally different from that in H5R and 409R, suggesting possible different genetic background difference behind their resistance.

Fine Mapping Revealed at Least Two Resistance Loci Presented in DingWen

To investigate the DW inheritance of the resistance, a segregating population was constructed using a susceptible material HZSX crossed with DW. Parental lines (DW and HZSX) and 139 F_1 individuals were subsequently inoculated with the *PbXm* and *scored* for their resistance. As shown in **Table 1**, DW showed complete resistance with DI of 0, whereas HZSX was completely susceptible with DI of 100% (**Table 1**). The F_1 plants from a cross of DW with HZSX exhibited roughly 1:1 segregation (among 139 F_1 individuals, 69 and 70 were resistant and susceptible, respectively), indicating that the CR loci in DW are controlled by a single dominant gene (**Figure 2A** and **Table 1**).

In order to identify the resistance loci, BSA (bulked segregation analysis) pools consisting of R- and S-plants were used for high-throughput sequencing. A total of 278,237,232 and 271,955,410 raw reads were obtained from R- and S-pools, respectively (**Supplementary Table 2**). After removing the adaptor sequence and filtering off low-quality reads, we recovered over 93% clean reads from the R- and S-pools, respectively, with Q30 over 87% (**Supplementary Table 2**). The clean reads were mapped to *B. rapa* reference genome (V3.0), and the total mapping rates were over 96%.

According to the ED of SNP and InDel (**Figure 2B**), a region of about 15 Mb was identified on A08 chromosome associating with CR (**Figure 2B**). To narrow down the region

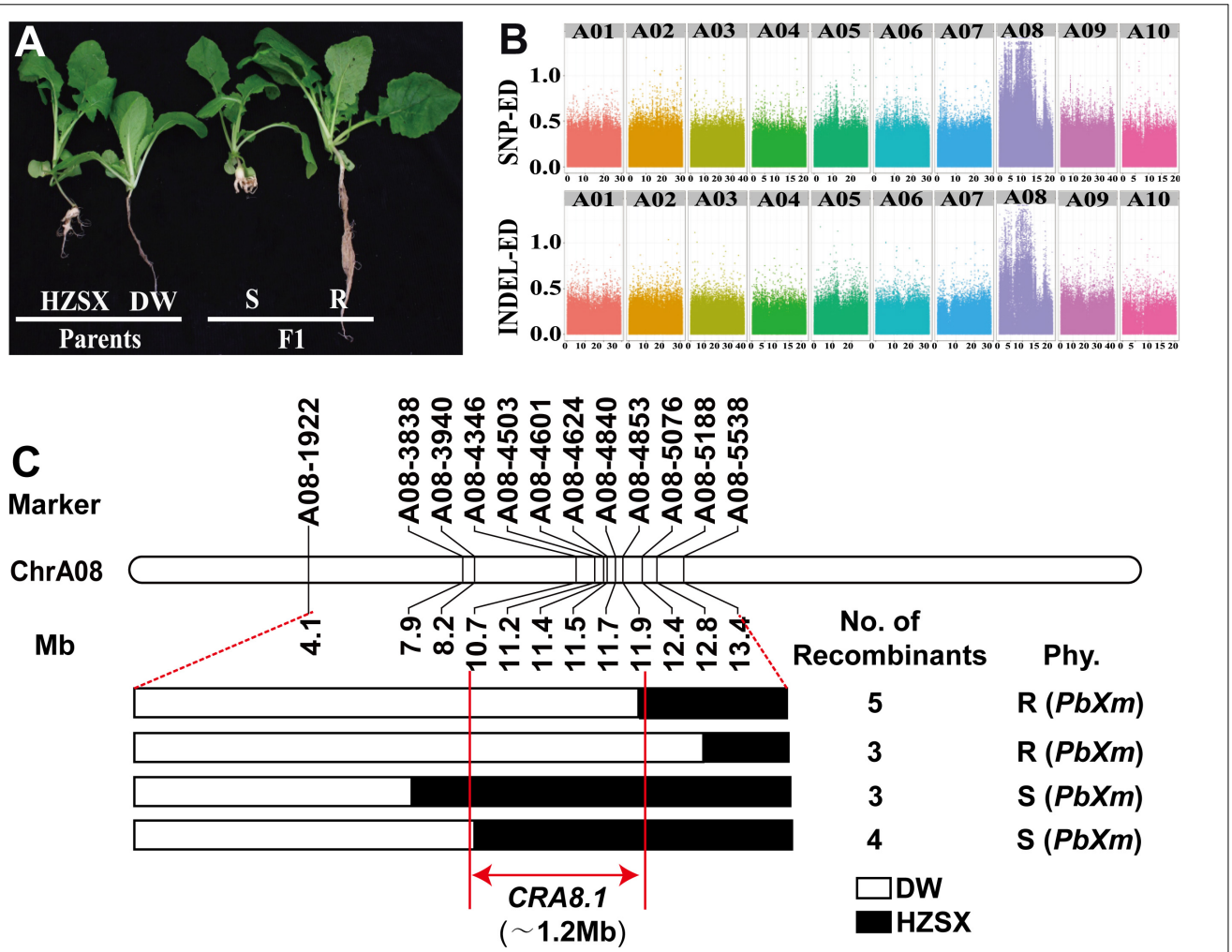


FIGURE 2 | Primary mapping of *CRA8.1* in *B. rapa*. **(A)** Root phenotypes of DW, HZSX, and F_1 after inoculation with *PbXm*. S, susceptible; R, resistant. **(B)** The Euclidean distance (ED) values of SNPs and InDels on chromosomes from the A genome. **(C)** Fine mapping of *CRA8.1* region with InDel markers. Fifteen recombinants were identified from F_1 population from markers A08-1922 to A08-5538. Black or white represents the regions found in HZSX or DW, respectively. Red lines indicate the *CRA8.1* locus (1.2 Mb) identified from this study.

in genome, 12 molecule markers from the developed molecular markers were screened, and 15 recombinants were screened between markers A08-1922 and A08-5538 in F_1 population. They were divided into four groups carrying different chromosomal fragments based on genotyping results (**Figure 2C**). Based on the resistance phenotype and the genotyping results, the CR loci were primarily mapped between A08-4346 and A08-4853 intervals, corresponding to a region of 1.2 Mb on chromosome A08 of *B. rapa* genome (V3.0) (**Figure 2C**). Interestingly, a previously known resistance locus *PbBa8.1* overlaps with this region (Zhan et al., 2020). Since DW was highly resistant to *PbXm* pathogen, while H5R was not, this is a new locus conferring different CR profiles from *PbBa8.1*, and we named this locus as *CRA8.1* hereafter.

To further narrow down the region of *CRA8.1* and eventually identify the candidate genes, we expanded the F_1 population for fine mapping. Finally, 39 recombinants were obtained from 3290 F_1 plants. Two different recombinant individuals, namely, WHR10 and CW75, were obtained, and they were sterile. We backcrossed them with HZSX, and the progenies were inoculated by *PbZj*, *PbXm*, and *PbXm*$^+$, together with DW and Chinese cabbage 91-12, as positive and negative controls, respectively (**Figure 3**). When *PbXm* or *PbZj* was used, the backcross progeny of these two recombinants with HZSX also showed 1:1 segregation (**Figures 3B–E** and **Supplementary Table 3**); however, when *PbXm*$^+$ was used, the offspring of WHR10 showed a segregation ratio of 1:1, while those from CW75 were all susceptible (**Figures 3F,G** and **Supplementary Table 3**). Based on the BSA sequencing, the fragment between A08-4346 and A08-4624 intervals, in a total length of 765 kb, might be responsible for the resistance phenotype variation between CW75 and WHR10 for *PbXm*$^+$ (**Figure 3A**). Therefore, we concluded that *CRA8.1* contains at least two disease-resistant loci, conferring resistance to different isolates (*PbXm*$^+$ and *PbXm*/*PbZj*). The

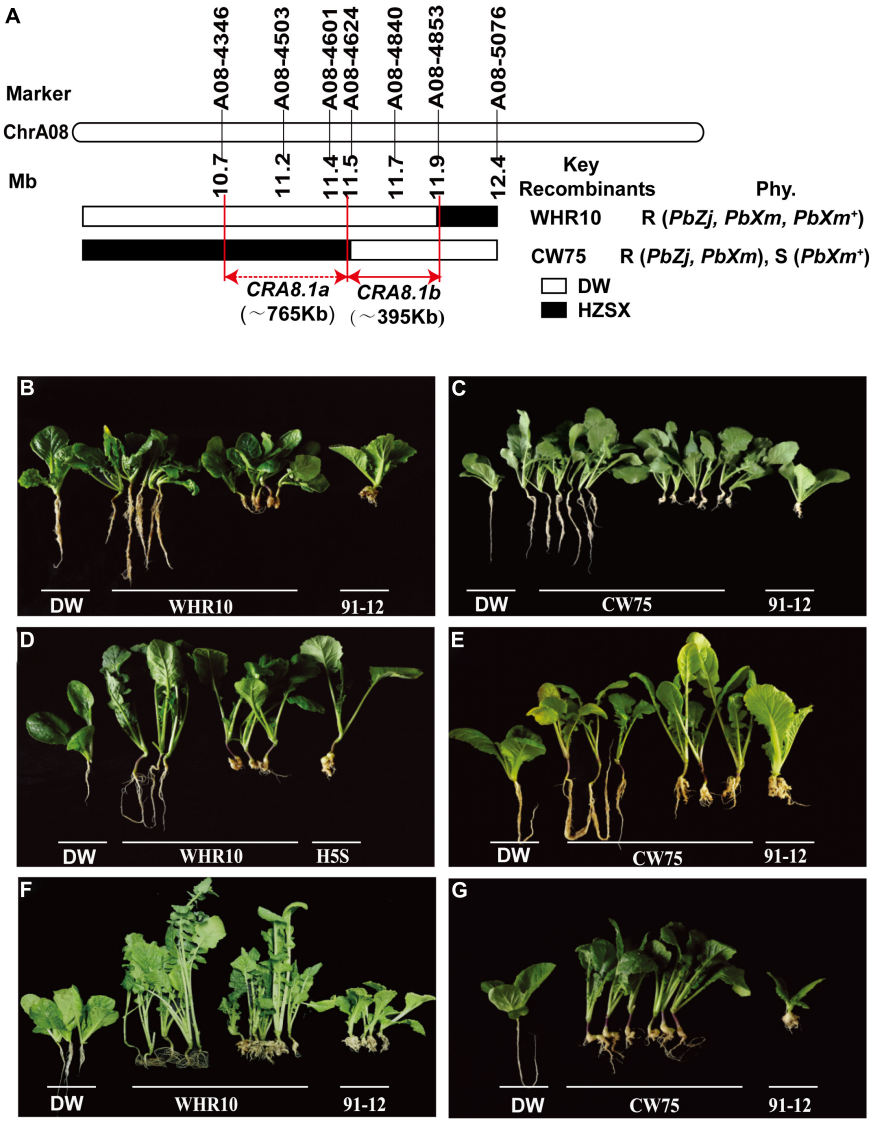


FIGURE 3 | *CRA8.1a* or *CRA8.1b* loci identified by fine mapping from A08-4346 to A08-4853 markers. **(A)** Genotyping of two representative recombinants CW75 and WHR10, carrying *CRA8.1b* or *CRA8.1a* and *CRA8.1b*, respectively. Disease phenotype of WHR10 and CW75 progeny plants after inoculation of *PbZj* **(B,C)**, *PbXm* **(D,E)**, or *PbXm*$^{+}$ **(F,G)**, respectively.

candidate region named as *CRA8.1b* covers a region of 395 kb between markers A08-4624 and A08-4853, and it contributes to resistance to *PbZj*/*PbXm* isolates. The other locus *CRA8.1a*, corresponding to a region between A08-4346 and A08-4624, may confer resistance to *PbXm*$^{+}$ by itself or coordinately with *CRA8.1b* (**Figure 3A**).

Three Candidate Clubroot-Resistant Genes Identified in *CRA8.1* Region Based on Resequencing and Sequence Comparison

In order to identify candidate CR genes in the region between A08-4346 and A08-4853 intervals on A08 chromosome, we resequenced one of the resistant recombinant individuals R59 generated from the F_1 population. The sequencing yielded 422 million clean reads with 63 billion clean bases. We manually assembled this region using ECD04 as the reference genome (unpublished) and compared gene distribution between DW, HZSX, and, a widely used *B. napus* variety, ZS11 (**Figure 4**). The total length of this assembly was 1.4 Mb (including the fragment between markers A08-4346 and A08-5076), and we found a good collinearity between DW, Z1 (HZSX), ZS11, and ECD04 (**Figure 4A**). Pairwise comparison within this segment suggested there were on average level of 14,395 SNPs and 2776 InDels that were different in DW to ZS11, ECD04, or Z1 varieties (**Figure 4C**).

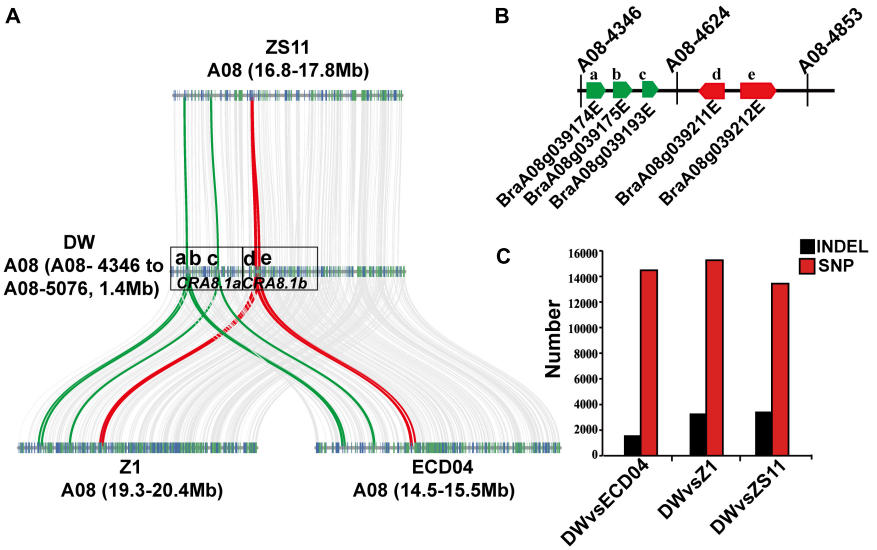


FIGURE 4 | Genome structure comparison between homologous regions on A08 chromosome from DW, HZSX (Z1), ECD04, and ZS11. **(A)** Gene synteny analysis between chromosomal regions from A08-4346 to A08-5076 on A08 chromosome from DW, ECD04, ZS11, and Z1. a–e represent five candidate R genes found within this region. a–c belongs to the *CRA8.1a* region. d, e belongs to the *CRA8.1b* region. **(B)** Genomic orientation of the five candidate R genes between A08-4346 and A08-4853 on A08 chromosome. Green and red bars represent the candidate genes in *CRA8.1a* and *CRA8.1b* loci, respectively. **(C)** Number of SNPs and InDels in DW compared with ECD04, Z1, and ZS11 between the target regions flanked by markers of A08-4346 and A08-5076.

It is reported that 61% of the cloned R genes code for NLRs and 19% of R genes for RLPs/RLKs (Kourelis and van der Hoorn, 2018). Based on NLR/RLP annotation, we identified five R genes within the candidate region of A08 chromosome, corresponding to the following gene sequences of ECD04: *CRA8.1.1* (*BraA08g039174E*), *CRA8.1.2* (*BraA08g039175E*), *CRA8.1.3* (*BraA08g039193E*), *CRA8.1.4* (*BraA08g039211E*), and *CRA8.1.5* (*BraA08g039212E*) (**Figure 4A** and **Table 2**). Among those candidate genes, the homologs of *BraA08g039174E*, *BraA08g039175E,* and *BraA08g039193E* are typical RLP genes and homologs of *BraA08g039211E* and *BraA08g039212E* are TIR-NBS-LRR genes (**Figure 4B** and **Table 2**). Homologs of *BraA08g039174E*, *BraA08g039175E,* and *BraA08g039193E* are located in *CRA8.1a* region, while the remaining two are in *CRA8.1b* region (**Figure 4A**).

We reason that if an R gene is present in both resistant and susceptible plants, the difference may reside on gene sequence (including CDS and promoter) or expression level, or both. Since DW is a heterozygous material for resistance, in order to find out the resistant allelic genes in DW, we designed primers covering 3.0 kb upstream and the whole CDS region using information of genome resequencing (**Supplementary Table 1**). After T-A cloning and sequencing, genes with sequence variation between DW and HZSX were considered as genes possibly responsible for the new type of CR in DW.

Then, in order to clarify whether there is a functional diversification between resistant versus susceptible varieties, we extracted gene homologs from ECD04, ZS11, SL (Sheng Li), and QU (Quinta),[7] the latter three as susceptible *B. napus* representatives. We compared gene coding sequences using BioEdit software (**Supplementary Figures 1–5**). We found that the protein sequences of *BraA08g039174E* and *BraA08g039175E*, both as LRR-domain-only proteins, were similar between different accessions (**Supplementary Figures 1**, **2**). As for the third LRR-domain-only protein, *BraA08g039193E*, the DW and ECD04 protein sequences were quite different compared with the ones in ZS11, SL, and QU materials (**Supplementary Figure 3**). The protein sequence of *BraA08g039211E* in DW showed two missing amino acids (S223, Y224 in ECD04, and Z1) and one substitution (E191) in TIR domain; otherwise, the sequences were quite similar (**Supplementary Figure 4**). In the case of *BraA08g039212E*, the sequence in DW was almost identical to that of ECD04, but different from that in susceptible Zl, ZS11, and other accessions (**Supplementary Figure 5**).

We also quantified the relative expression of these candidate genes in DW or Z1 (HZSX) upon inoculation of *PbXm*$^{+}$ or *PbZj*, except for *BraA08g039175E* whose expression was too low (**Figure 5**). As shown in **Figure 5**, the expression of *BraA08g039174E* was quite similar between DW and

TABLE 2 | List of CR candidate genes identified within *CRA8.1* loci.

Gene ID (ECD04)	AGI (Arabidopsis gene)	Gene type
BraA08g039174E	AT3G05360.1	RLP
BraA08g039175E	AT1G71400.1	RLP
BraA08g039193E	AT3G05360.1	RLP
BraA08g039211E	AT4G19500.1	TIR-NBS-LRR
BraA08g039212E	AT4G19510.1	TIR-NBS-LRR

[7]http://yanglab.hzau.edu.cn/BnTIR

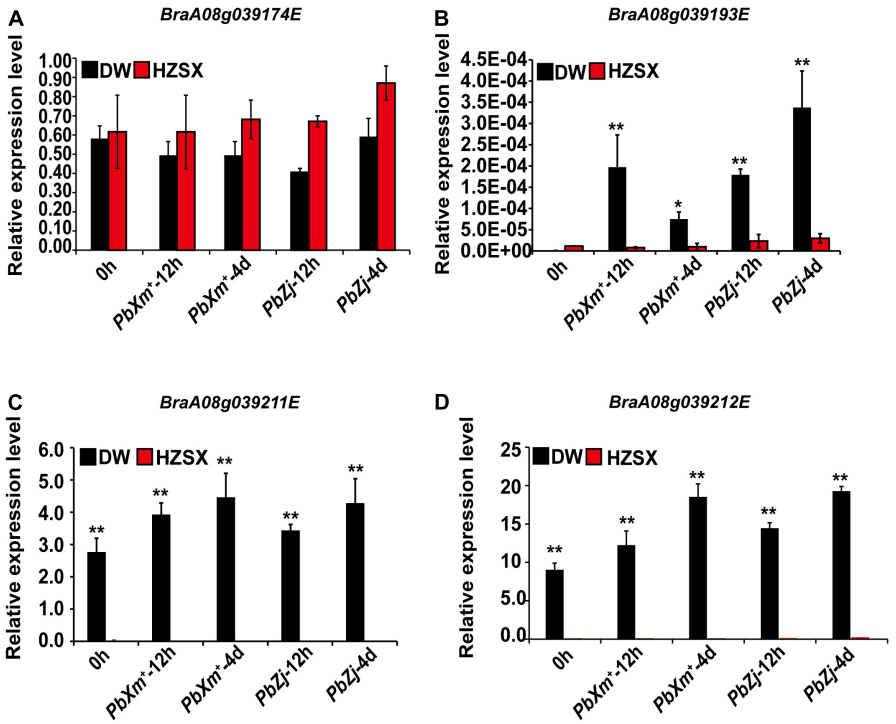


FIGURE 5 | Relative expression of candidate CR genes between DW (black) and HZSX (red) before and after inoculation of *PbXm*$^{+}$ and *PbZj* isolates. **(A–D)** Relative expression of homologs of *BraA08g039174E, BraA08g039193E*, *BraA08g039211E*, and *BraA08g039212E*, respectively. ** represents $p \leq 0.01$ by Student's *t*-test.

HZSX, both before and after inoculation (**Figure 5A**). On the contrary, we observed a clear difference on the expression levels of *BraA08g039211E* and *BraA08g039212E* between DW and HZSX, in particular an obvious induction on *BraA08g039193E* in DW upon infection (**Figures 5B–D**). Since these genes are NBS-LRR or RLP genes and also show differential expression between susceptible and resistant materials responding to pathogen infection, they are considered as the final candidate R genes awaiting further functional verification.

In order to find the reason for different transcript levels between resistant and susceptible materials, especially in DW, we extracted promoter sequence and predicted regulatory *cis*-elements within these regions from DW, ECD04, ZS11, and Z1(HZSX) (**Figure 6**). As shown in **Figure 6**, similar patterns were observed in promoter regions of *BraA08g039193E*, *BraA08g039211E,* and *BraA08g039212E* from ECD04 and their corresponding orthologs in DW, but different from that in susceptible ZS11 and Z1 (**Figure 6**). For example, the MeJA-responsive elements in the promoter regions of *BraA08g039211E* and *BraA08g039212E* were very similar in DW and ECD04, but not compared to Z1 and ZS11 (**Figures 6A,B**). In the case of *BraA08g039193E*, the *cis*-elements in ZS11 and Z1 were similar, followed by DW, but with additional defense- and stress-responsive elements. The *cis*-element in ECD04 was different from the remaining three, and whether this is of any biological relevance needs further verification (**Figure 6C**).

Taken together, our data from the expression profile, protein sequence, and promoter *cis*-element analysis suggest that *BraA08g039211E* and *BraA08g039212E* from *CRA8.1b* region are the most likely candidates leading to resistance to *PbZj*/*PbXm*. In addition, *BraA08g039193E* is likely to be the main candidate gene in *CRA8.1a*, but *BraA08g039211E* and *BraA08g039212E* may interact with *BraA08g039193E* to resist *PbXm*$^{+}$ isolate coordinately.

DISCUSSION

In this study, we discovered a unique *B. rapa* material DW that showed different resistance profiles other than H5R (*PbBa8.1*) or Huayouza 62R (*CRb*), suggesting possible new CR genes involved in its resistance. Fine mapping narrowed down to a genomic region with two NBS-LRR genes and one RLP gene as potential candidates. Once validated for their function as CR genes, those candidate genes could confer broader and higher resistance to more virulent clubroot pathogens. Therefore, genes from DW provide a new gene resource for CR breeding in future.

CRA8.1 Is Different From *PbBa8.1* and Confers Superior Resistance to Diverse Pathogens

Numerous CR loci have been identified previously from ECD04 and other *B. rapa* materials, for instance, *PbBa8.1*

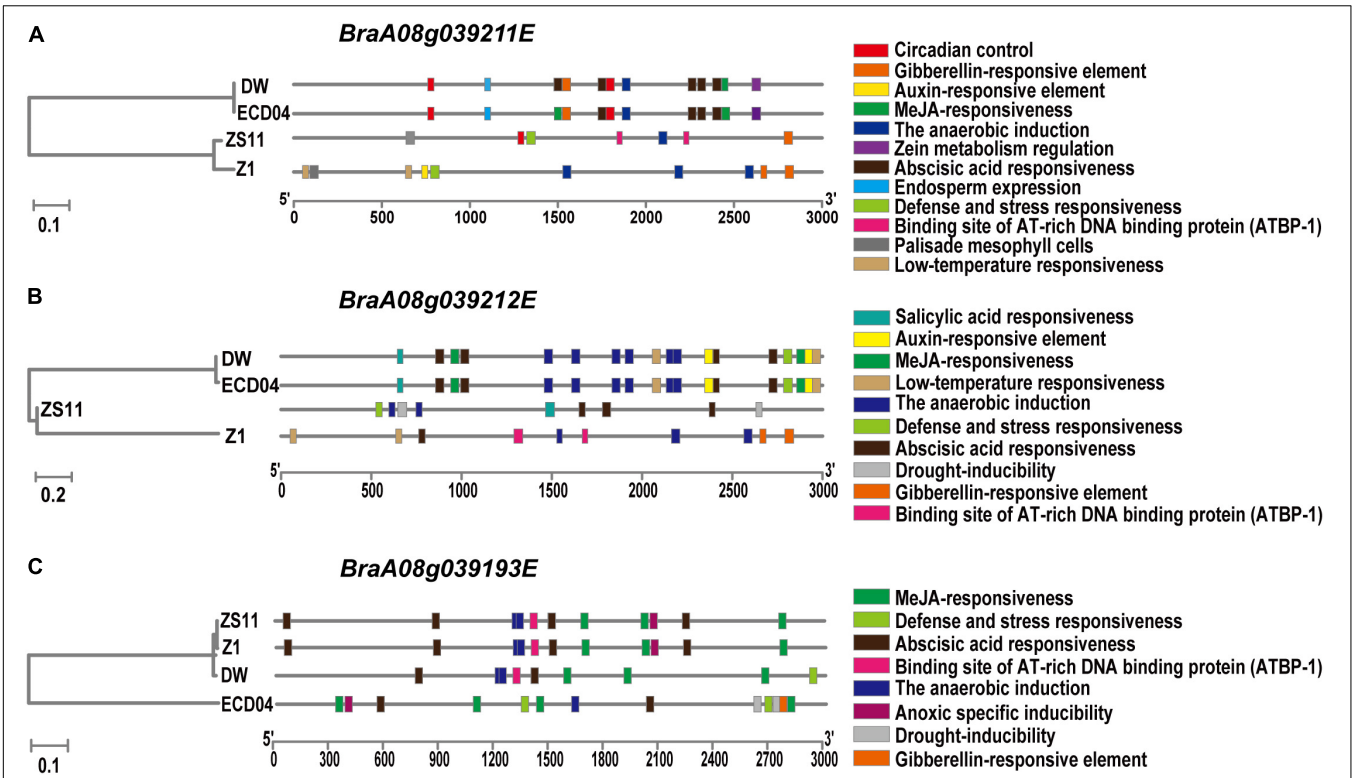


FIGURE 6 | Analysis of *cis*-acting elements in promoter region of the homologs of *BraA08g039211E* **(A)**, *BraA08g039212E* **(B)**, and *BraA08g039193E* **(C)** between resistant (DW and ECD04) and susceptible (ZS11 and Z1/HZSX) materials.

and *Crr1a* on A08 chromosome (Chen et al., 2013; Hatakeyama et al., 2013) (**Table 3**). In this study, our locus *CRA8.1* is also located on chromosome A08 in DW, and genetically, this locus is dominant based on the F_1 population segregation results. *PbBa8.1* is the source of resistance in H5R; however, H5R was susceptible to *PbXm* pathogen, suggesting new genes in DW being responsible for its resistance. In the process of fine mapping, we found that there were two loci *CRA8.1a* and *CRA8.1b* in the primary localization interval of *CRA8.1*. Homologs of *BraA08g039211E* and *BraA08g039212E* are the most promising candidate genes in DW for *CRA8.1b*, which may play an important role against *PbZj* and *PbXm* isolates. However, what is the different resistance of *BraA08g039211E* and *BraA08g039212E* in DW and H5R remains to be further studied. The resistance of DW to $PbXm^+$ comes from *CRA8.1a* itself or *CRA8.1a* and *CRA8.1b* together. So, it is not clear whether there is an interaction between the homologous genes of *BraA08g039211E* and *BraA08g039212E* and other resistance genes, or there are other types of resistance genes independently involved in the resistance response. In order to solve this problem, it is necessary to deepen the fine mapping of *CRA8.1* loci, which involves the screening and isolation of recombinant plants with *CRA8.1a* locus only and subsequent inoculation of its offspring with $PbXm^+$ to observe the segregation ratio.

The New Type of Clubroot-Resistant Genes That Can Be Used for Revealing the Novel Resistance Mechanisms

Based on gene distribution around *CRA8.1* locus in ECD04, ZS11, DW and HZSX (Z1), the resistance spectrum of DW is unique and most likely coming from the *CRA8.1a* fragment (**Figure 4**). The candidate genes were the only NBS-LRR or RLP/RLK genes within this region, of which five can be quantified for relative expression. Among the three final candidates, *BraA08g039211E* and *BraA08g039212E* are actually located head to head adjacently on the chromosome, and both are drastically induced from 12 h up to 4 days in DW after clubroot infection (**Figures 4B**, **5C**,**D**). Previous studies have shown that some of the R genes work in pair, and they are physically located very close on the chromosome, such as TNL pair genes RRS1/RPS4 from *Arabidopsis thaliana* (Le Roux et al., 2015) and CNL pair genes RGA5/RGA4 in rice (Cesari et al., 2013). So, *BraA08g039211E* and *BraA08g039212E* as TIR-NBS-LRR genes might be very important for host response to the clubroot pathogen. The third gene, *BraA08g039193E* is a RLP protein. This is distinct from the previously cloned NLR-type genes and mediates immune responses at the PTI level. Once the candidate gene is validated, what is the disease resistance mechanism is also a very interesting scientific question. In addition, *BraA08g039211E* has little residue difference between DW and HZSX, *BraA08g039193E* and *BraA08g039212E*, the sequence in DW was quite different

TABLE 3 | Clubroot-resistant loci presently known in *Brassica rapa*.

Location	Source	Name of loci	Pathogen	Classification	References
A01	G004 (Siloga derived)	*Crr2*	Wakayama-01	Williams (race 4)	Suwabe et al., 2003
A01	ECD04	*PbBa1.1*	Pb2, Pb7	Williams (race 2, race 7)	Chen et al., 2013
A02	C9 (Debra derived)	*CRc*	K04	n.a.[a]	Sakamoto et al., 2008
A02	T19 (Pluto derived)	*Rcr8*	Pathotype 5x	Williams (race 5[b])	Yu et al., 2017
A03	ECD04	*PbBa3.1*	Pb2	Williams (race 2)	Chen et al., 2013
A03	ECD04	*PbBa3.2*	Pb10	Williams (race 10)	Chen et al., 2013
A03	ECD04	*PbBa3.3*	Pb7	Williams (race 7)	Chen et al., 2013
A03	Chinese cabbage inbred line 85-74	*CRd*	Pb4	Williams (race 4)	Pang et al., 2018
A03	N-WMR-3 (Milan White derived)	*Crr3*	Ano-01	Williams (race 4)	Hirai et al., 2004
A03	K10 (CR Kanko derived)	*CRk*	M85 and K04	n.a.[a]	Sakamoto et al., 2008
A03	Chinese cabbage cultivar, CR Shinki	*CRb*	Pb4	Williams (race 4)	Piao et al., 2004
A03	T19 (Pluto derived)	*Rcr4*	Pathotypes 2, 3, 5, 6, and 8	Williams	Yu et al., 2017
A03	ECD1, ECD2, ECD4	*BraA.CR.a*	n.a.	n.a	Hirani et al., 2018
A03	ECD3	*BraA.CR.c*	n.a.	n.a	Hirani et al., 2018
A03	ECD02	*BraA3P5X.CRa/b^{Kato}1.1*	Pathotype 5x	Williams (race 5[b])	Fredua-Agyeman et al., 2020
A03	ECD02	*BraA3P5X.CRa/b^{Kato}1.2*	Pathotype 5x	Williams (race 5[b])	Fredua-Agyeman et al., 2020
A03	Pak choy cv. 'Flower Nabana'	*Rcr1*	Pathotype 3	Williams (race 3)	Chu et al., 2014
A03	Chinese cabbage inbred line T 136-8	*CRa*	Pathotype 2	Williams (race 2)	Matsumoto et al., 1998
A03	Chinese cabbage cv. 'Jazz'	*Rcr2*	Pathotype 3	Williams (race 3)	Huang et al., 2017
A06	G004 (Siloga derived)	*Crr4*	Wakayama-01, Ano-01	Williams (race 4)	Suwabe et al., 2003
A08	G004 (Siloga derived)	*Crr1*	Wakayama-01	Williams (race 4)	Suwabe et al., 2003
A08	ECD1, ECD2, ECD3, ECD4	*BraA.CR.b*	n.a.	n.a	Hirani et al., 2018
A08	ECD04	*PbBa8.1*	Pb4	Williams (race 4)	Chen et al., 2013
A08	96-6990-2 (Waaslander derived)	*Rcr3*	Pathotype 3H	Williams (race 3)	Karim et al., 2020
A08	T19 (Pluto derived)	*Rcr9*	Pathotype 5x	Williams (race 5[b])	Yu et al., 2017
A08	ECD04	*Rcr9^{Wa}*	Pathotype 5x	Williams (race 5[b])	Karim et al., 2020

[a]*n.a., not available.* [b]*Pathotype 5x is a new pathotype identified by Yu et al. (2017).*

compared with the ones in ZS11, SL, and QU materials, so they can be used as good materials for resistance mechanism research in future.

DingWen Is an Excellent Clubroot-Resistant Donor Plant for Resistance Improvement of *Brassica napus*

As mentioned above, clubroot is a fast-spreading soil-borne disease which calls for fast and effective resistant varieties. The current resistance cultivars can revert to susceptible upon long terms of use, and single or a handful of CR genes cannot assure resistance to different pathotypes of the clubroot pathogens. Also, the accumulation of mutations eventually would result in the loss of resistance. For example, in Alberta of Canada, a "new" brassicae pathogenic type that can overcome resistance was discovered only 4 years after the CR variety release (Strelkov et al., 2016; Fredua-Agyeman et al., 2018). Most CR genes used currently are coming from *B. rapa*, since crossability between *B. rapa* and *B. napus* is very good, and CR genes from *B. rapa* can readily be transferred to *B. napus*. The current two widely used CR varieties of *B. napus*, namely, H5R and 409R, are resistant to *PbZj* but not to *PbTc* and *PbXm* (Shah et al., 2020). Previous studies have shown that different disease resistance genes' aggregation can improve the resistance to clubroot (Suwabe et al., 2006; Shah et al., 2020). In this study, the *CRA8.1* locus located in DW contains at least two CR loci, which is superior to *PbBa8.1* and *CRb* on A08 chromosome, so the two resistance genes of *CRA8.1* can be introduced into excellent *B. napus* parents to improve the CR in breeding, which can save cost and improve work efficiency.

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

YW performed the experiments, analyzed the data, prepared the figures, and drafted the manuscript. XX helped to analyze the data. CZ and PC conceived the study, participated in its coordination, and helped to draft the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant nos. U20A2034 and 31871659) and China Agriculture Research System (CARS-12) to CZ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.898108/full#supplementary-material

Supplementary Figure 1 | Protein sequence alignments of homologs of *BraA08g039174E* from DW, ECD04, Z1, ZS11, SL, and QU. Red solid lines represent the LRR domain.

Supplementary Figure 2 | Protein sequence alignments of homologs of *BraA08g039175E* from DW, ECD04, Z1, ZS11, SL, and QU. Red solid lines represent the LRR domain.

Supplementary Figure 3 | Protein sequence alignments of homologs of *BraA08g039193E* from DW, ECD04, Z1, ZS11, SL, and QU. Red solid lines represent the LRR domain.

Supplementary Figure 4 | Protein sequence alignments of homologs of *BraA08g039211E* from DW, ECD04, Z1, ZS11, SL, and QU. Green solid lines represent the TIR domain, orange solid lines represent the NB-ARC domain, and red solid lines represent the LRR domain. The symbol "**" indicates the two missing residues in DW compared with other accessions. The symbol "*" represents the specific residues in DW.

Supplementary Figure 5 | Protein sequence alignments of homologs of *BraA08g039212E* from DW, ECD04, Z1, ZS11, SL, and QU. Green solid lines represent the TIR domain, orange solid lines represent the NB-ARC domain, and red solid lines represent the LRR domain. The symbol "*" represents the specific residues in DW.

Supplementary Table 1 | Primers used in this study.

Supplementary Table 2 | Quality control of BSA sequencing data.

Supplementary Table 3 | Phenotype statistics of clubroot resistance for progeny of WHR10 and CW75 recombinants.

REFERENCES

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, 879–887. doi: 10.1038/s41477-018-0289-4

Buczacki, S. T., Toxopeus, H., Mattusch, P., Johnston, T. D., Dixon, G. R., and Hobolth, L. A. (1975). Study of physiologic specialization in Plasmodiophora brassicae : proposals for attempted rationalization through an international approach. *Transac. Brit. Mycol. Soc.* 65, 295–303. doi: 10.1016/S0007-1536(75)80013-1

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* 25, 1463–1481. doi: 10.1105/tpc.112.107201

Chai, A. L., Xie, X. W., Shi, Y. X., and Li, B. J. (2014). Research status of clubroot (Plasmodiophora brassicae) on cruciferous crops in China. *Can. J. Plant Pathol.* 36, 142–153. doi: 10.1080/07060661.2013.868829

Chen, J., Jing, J., Zhan, Z., Zhang, T., Zhang, C., and Piao, Z. (2013). Identification of novel QTLs for isolate-specific partial resistance to Plasmodiophora brassicae in Brassica rapa. *PLoS One* 8:e85307. doi: 10.1371/journal.pone.0085307

Chen, J., Pang, W., Chen, B., Zhang, C., and Piao, Z. (2015). Transcriptome Analysis of Brassica rapa Near-Isogenic Lines Carrying Clubroot-Resistant and -Susceptible Alleles in Response to Plasmodiophora brassicae during Early Infection. *Front. Plant Sci.* 6:1183. doi: 10.3389/fpls.2015.01183

Chu, M., Song, T., Falk, K. C., Zhang, X., Liu, X., Chang, A., et al. (2014). Fine mapping of Rcr1 and analyses of its effect on transcriptome patterns during infection by Plasmodiophora brassicae. *BMC Genom.* 15:1166. doi: 10.1186/1471-2164-15-1166

Deng, Y., Ning, Y., Yang, D. L., Zhai, K., Wang, G. L., and He, Z. (2020). Molecular basis of disease resistance and perspectives on breeding strategies for resistance improvement in crops. *Mol. Plant* 13, 1402–1419. doi: 10.1016/j.molp.2020.09.018

Dixon, G. R. (2009). The occurrence and economic impact of Plasmodiophora brassicae and clubroot disease. *J. Plant Grow. Regul.* 28, 194–202. doi: 10.1007/s00344-009-9090-y

Elke, D., Martin, F., Enrico, G. A. L., Katsunori, H., and Masashi, H. (2009). Status and perspectives of clubroot resistance breeding in crucifer crops. *J. Plant Growth Regul.* 28, 265–281. doi: 10.1007/s00344-009-9100-0

Fredua-Agyeman, R., Hwang, S. F., Strelkov, S. E., Zhou, Q., and Feindel, D. (2018). Potential loss of clubroot resistance genes from donor parent Brassica rapa subsp rapifera (ECD04) during doubled haploid production. *Plant Pathol.* 67, 892–901. doi: 10.1111/ppa.12816

Fredua-Agyeman, R., Jiang, J., Hwang, S. F., and Strelkov, S. E. (2020). QTL Mapping and Inheritance of Clubroot Resistance Genes Derived From Brassica rapa subsp. rapifera (ECD02) Reveals Resistance Loci and Distorted Segregation Ratios in Two F2 Populations of Different Crosses. *Front. Plant Sci.* 11:899. doi: 10.3389/fpls.2020.00899

Gahatraj, S., Shrestha, S. M., Devkota, T. R., and Rai, H. H. (2019). A review on clubroot of crucifers: symptoms, life-cycle of pathogen, factors affecting severity, and management strategies. *Arch. Agricul. Environ. Sci.* 4, 342–349. doi: 10.26832/24566632.2019.0403012

Grund, E., Tremousaygue, D., and Deslandes, L. (2019). Plant NLRs with integrated domains: unity makes strength. *Plant Physiol.* 179, 1227–1235. doi: 10.1104/pp.18.01134

Hatakeyama, K., Suwabe, K., Tomita, R. N., Kato, T., Nunome, T., Fukuoka, H., et al. (2013). Identification and characterization of Crr1a, a gene for resistance to clubroot disease (Plasmodiophora brassicae Woronin) in Brassica rapa L. *PLoS One* 8:e54745. doi: 10.1371/journal.pone.0054745

Hirai, M., Harada, T., Kubo, N., Tsukada, M., Suwabe, K., and Matsumoto, S. (2004). A novel locus for clubroot resistance in Brassica rapa and its linkage markers. *Theor. Appl. Genet.* 108, 639–643. doi: 10.1007/s00122-003-1475-x

Hirani, A. H., Gao, F., Liu, J., Fu, G., Wu, C., McVetty, P., et al. (2018). Combinations of independent dominant loci conferring clubroot resistance in all four turnip accessions (Brassica rapa) from the European clubroot differential set. *Front. Plant Sci.* 9:1628. doi: 10.3389/fpls.2018.01628

Huang, Z., Peng, G., Liu, X., Deora, A., Falk, K. C., Gossen, B. D., et al. (2017). Fine Mapping of a Clubroot Resistance Gene in Chinese Cabbage Using SNP Markers Identified from Bulked Segregant RNA Sequencing. *Front Plant Sci.* 8:1448. doi: 10.3389/fpls.2017.01448

Hwang, S. F., Strelkov, S. E., Feng, J., Gossen, B. D., and Howard, R. J. (2012). Plasmodiophora brassicae: a review of an emerging pathogen of the Canadian canola (Brassica napus) crop. *Mol. Plant Pathol.* 13, 105–113. doi: 10.1111/j.1364-3703.2011.00729.x

Karim, M. M., Dakouri, A., Zhang, Y., Chen, Q., Peng, G., Strelkov, S. E., et al. (2020). Two Clubroot-Resistance genes. *Rcr3 and Rcr9wa, mapped in Brassica rapa using bulk segregant RNA sequencing. Int. J. Mol. Sci.* 21:5033. doi: 10.3390/ijms21145033

Kim, S. H., Qi, D., Ashfield, T., Helm, M., and Innes, R. W. (2016). Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* 351, 684–687. doi: 10.1126/science.aad3436

Kim, W., Prokchorchik, M., Tian, Y., Kim, S., Jeon, H., and Segonzac, C. (2020). Perception of unrelated microbe-associated molecular patterns triggers conserved yet variable physiological and transcriptional changes in Brassica rapa ssp. *Pekinensis. Hortic Res.* 7:186. doi: 10.1038/s41438-020-00410-0

Kourelis, J., and van der Hoorn, R. A. L. (2018). Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579

Larkan, N. J., Ma, L., and Borhan, M. H. (2015). The Brassica napus receptor-like protein RLM2 is encoded by a second allele of the LepR3/Rlm2 blackleg resistance locus. *Plant Biotechnol. J.* 13, 983–992. doi: 10.1111/pbi.12341

Larkan, N. J., Ma, L., Haddadi, P., Buchwaldt, M., Parkin, I., Djavaheri, M., et al. (2020). The Brassica napus wall-associated kinase-like (WAKL) gene Rlm9 provides race-specific blackleg resistance. *Plant J.* 104, 892–900. doi: 10.1111/tpj.14966

Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Tremousaygue, D., Kraut, A., et al. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161, 1074–1088. doi: 10.1016/j.cell.2015.04.025

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, Q., Shah, N., Zhou, X., Wang, H., Yu, W., Luo, J., et al. (2021). Identification of Micro Ribonucleic Acids and their Targets in Response to Plasmodiophora brassicae Infection in Brassica napus. *Front. Plant Sci.* 12:734419. doi: 10.3389/fpls.2021.734419

Matsumoto, E., Yasui, C., Ohi, M., and Tsukada, M. (1998). Linkage analysis of RFLP markers for clubroot resistance and pigmentation in Chinese cabbage (Brassica rapa ssp. *Pekinensis). Euphytica* 104, 79–86. doi: 10.1023/A:1018370418201

Mehraj, H., Akter, A., Miyaji, N., Miyazaki, J., Shea, D. J., Fujimoto, R., et al. (2020). Genetics of clubroot and fusarium wilt disease resistance in brassica vegetables: the application of marker assisted breeding for disease resistance. *Plants* 9:726. doi: 10.3390/plants9060726

Nadil, S., Jincai, S., Shaowei, Y., Zhaochun, Y., Zuo, W., Fan, H., et al. (2019). Genetic variation analysis of field isolates of clubroot and their responses to Brassica napus lines containing resistant genes CRb and PbBa8.1 and their combination in homozygous and heterozygous state. *Mol. Breeding* 39:153. doi: 10.1007/s11032-019-1075-3

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCF tools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32, 1749–1751. doi: 10.1093/bioinformatics/btw044

Neik, T. X., Barbetti, M. J., and Batley, J. (2017). Current status and challenges in identifying disease resistance genes in brassica napus. *Front. Plant Sci.* 8:1788. doi: 10.3389/fpls.2017.01788

Nguyen, M. L., Monakhos, G. F., Komakhin, R. A., and Monakhos, S. G. (2018). The New Clubroot Resistance Locus is Located on Chromosome A05 in Chinese Cabbage (Brassica rapa L.). *Russ. J. Genet.* 54, 296–304. doi: 10.1134/S1022795418030080

Pang, W., Fu, P., Li, X., Zhan, Z., Yu, S., and Piao, Z. (2018). Identification and Mapping of the Clubroot Resistance Gene CRd in Chinese Cabbage (Brassica rapa ssp. Pekinensis). *Front. Plant Sci.* 9:653. doi: 10.3389/fpls.2018.00653

Pang, W., Liang, Y., Zhan, Z., Li, X., and Piao, Z. (2020). Development of a sinitic clubroot differential set for the pathotype classification of Plasmodiophora brassicae. *Front. Plant Sci.* 11:568771. doi: 10.3389/fpls.2020.568771

Piao, Z. Y., Deng, Y. Q., Choi, S. R., Park, Y. J., and Lim, Y. P. (2004). SCAR and CAPS mapping of CRb, a gene conferring resistance to Plasmodiophora brassicae in Chinese cabbage (Brassica rapa ssp pekinensis). *Theor. Appl. Genet.* 108, 1458–1465. doi: 10.1007/s00122-003-1577-5

Quinlan, A. R., and Hall, I. M. (2010). BED Tools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Sakamoto, K., Saito, A., Hayashida, N., Taguchi, G., and Matsumoto, E. (2008). Mapping of isolate-specific QTLs for clubroot resistance in Chinese cabbage (Brassica rapa L. Ssp pekinensis). *Theor. Appl. Genet.* 117, 759–767. doi: 10.1007/s00122-008-0817-0

Shah, N., Li, Q., Xu, Q., Liu, J., Huang, F., Zhan, Z., et al. (2020). CRb and PbBa8.1 Synergically Increases Resistant Genes Expression upon Infection of Plasmodiophora brassicae in Brassica napus. *Genes* 11:202. doi: 10.3390/genes11020202

Strelkov, S. E., Hwang, S., Manolii, V. P., Cao, T., and Feindel, D. (2016). Emergence of new virulence phenotypes of Plasmodiophora brassicae on canola (Brassica napus) in Alberta. *Canada. Eur. J. Plant Pathol.* 145, 517–529. doi: 10.1007/s10658-016-0888-8

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Fujimura, M., Nunome, T., et al. (2003). Identification of two loci for resistance to clubroot (Plasmodiophora brassicae Woronin) in Brassica rapa L. *Theor. Appl. Genet.* 107, 997–1002. doi: 10.1007/s00122-003-1309-x

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Kondo, M., Fujimura, M., et al. (2006). Simple sequence repeat-based comparative genomics between Brassica rapa and Arabidopsis thaliana : the genetic origin of clubroot resistance. *Genetics* 173, 309–319. doi: 10.1534/genetics.104.038968

Ueno, H., Matsumoto, E., Aruga, D., Kitagawa, S., Matsumura, H., and Hayashida, N. (2012). Molecular characterization of the CRa gene conferring clubroot resistance in Brassica rapa. *Plant Mol. Biol.* 80, 621–629. doi: 10.1007/s11103-012-9971-5

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, J., and Chai, J. (2020). Molecular actions of NLR immune receptors in plants and animals. Science China. *Life Sci.* 63, 1303–1316. doi: 10.1007/s11427-019-1687-6

Williams, P. H. (1966). A system for the determination of races of Plasmodiophora brassicae that infect Cabbage and Rutabaga. *Phytopathology* 56, 624–626.

Wu, Y., and Zhou, J. M. (2013). Receptor-like kinases in plant innate immunity. *J. Integr. Plant Biol.* 55, 1271–1286. doi: 10.1111/jipb.12123

Yu, F., Zhang, X., Peng, G., Falk, K. C., Strelkov, S. E., and Gossen, B. D. (2017). Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of Plasmodiophora brassicae in Brassica rapa. *Sci. Rep.* 7:4516. doi: 10.1038/s41598-017-04903-2

Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., et al. (2021). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 592, 105–109. doi: 10.1038/s41586-021-03316-6

Zhan, Z., Jiang, Y., Shah, N., Hou, Z., Zhou, Y., Dun, B., et al. (2020). Association of clubroot resistance locus PbBa8.1 with a linkage drag of high erucic acid content in the seed of the European turnip. *Front. Plant Sci.* 11:810. doi: 10.3389/fpls.2020.00810

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 in Plant Science

ORIGINAL RESEARCH
published: 15 December 2021
doi: 10.3389/fpls.2021.738805

Development and Validation of Kompetitive Allele-Specific PCR Assays for Erucic Acid Content in Indian Mustard [*Brassica juncea* (L.) Czern and Coss.]

Karanjot Singh Gill[1†], Gurpreet Kaur[1†], Gurdeep Kaur[1], Jasmeet Kaur[1], Simarjeet Kaur Sra[1], Kawalpreet Kaur[1], Kaur Gurpreet[1], Meha Sharma[1], Mitaly Bansal[2], Parveen Chhuneja[2] and Surinder S. Banga[1]*

[1] Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, [2] School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India

OPEN ACCESS

Edited by:
Harsh Raman,
New South Wales Department of Primary Industries, Australia

Reviewed by:
Anket Sharma,
University of Maryland, College Park, United States
Enrique Martinez Force,
Instituto de la Grasa (IG), Spain

***Correspondence:**
Gurpreet Kaur
gurpreetkaur@pau.edu
orcid.org/0000-0002-0660-9592

[†] These authors have contributed equally to this work

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

Received: 09 July 2021
Accepted: 20 October 2021
Published: 15 December 2021

Citation:
Gill KS, Kaur G, Kaur G, Kaur J, Kaur Sra S, Kaur K, Gurpreet K, Sharma M, Bansal M, Chhuneja P and Banga SS (2021) Development and Validation of Kompetitive Allele-Specific PCR Assays for Erucic Acid Content in Indian Mustard [Brassica juncea (L.) Czern and Coss.]. Front. Plant Sci. 12:738805. doi: 10.3389/fpls.2021.738805

Brassica juncea L. is the most widely cultivated oilseed crop in Indian subcontinent. Its seeds contain oil with very high concentration of erucic acid (≈50%). Of late, there is increasing emphasis on the development of low erucic acid varieties because of reported association of the consumption of high erucic acid oil with cardiac lipidosis. Erucic acid is synthesized from oleic acid by an elongation process involving two cycles of four sequential steps. Of which, the first step is catalyzed by β-ketoacyl-CoA synthase (KCS) encoded by the fatty acid elongase 1 (*FAE1*) gene in Brassica. Mutations in the coding region of the *FAE1* lead to the loss of KCS activity and consequently a drastic reduction of erucic acid in the seeds. Molecular markers have been developed on the basis of variation available in the coding or promoter region(s) of the *FAE1*. However, majority of these markers are not breeder friendly and are rarely used in the breeding programs. Present studies were planned to develop robust kompetitive allele-specific PCR (KASPar) assays with high throughput and economics of scale. We first cloned and sequenced *FAE1.1* and *FAE1.2* from high and low erucic acid (<2%) genotypes of *B. juncea* (AABB) and its progenitor species, *B. rapa* (AA) and *B. nigra* (BB). Sequence comparisons of *FAE1.1* and *FAE1.2* genes for low and high erucic acid genotypes revealed single nucleotide polymorphisms (SNPs) at 8 and 3 positions. Of these, three SNPs for *FAE1.1* and one SNPs for *FAE1.2* produced missense mutations, leading to amino acid modifications and inactivation of KCS enzyme. We used SNPs at positions 735 and 1,476 for genes *FAE1.1* and *FAE1.2*, respectively, to develop KASPar assays. These markers were validated on a collection of diverse genotypes and a segregating backcross progeny. KASPar assays developed in this study will be useful for marker-assisted breeding, as these can track recessive alleles in their heterozygous state with high reproducibility.

Keywords: KASPar, erucic acid, SNPs, *FAE* genes, allotetraploid

KEY MESSAGE

Developed KASPar assays for selection of low erucic acid are high throughput, robust, and codominant capable to track recessive alleles in their heterozygous state.

INTRODUCTION

Indian mustard [*Brassica juncea* (L.) Czern and Coss] is an important source of edible and industrial oil in the Indian subcontinent. Its seeds contain nearly 40% oil on a seed weight basis. Similar to other plants, oil in Indian mustard is a complex mixture of triglycerides (>95%) along with diacylglycerols (<5%), phytosterols, etc. Though mustard oil contains seven major fatty acids, erucic acid, a very long chain fatty acid (C22:1) is the predominant fatty acid contributing more than 50% to the total pool of fatty acids. Such levels are undesirable in view of their perceived role in cardiac lipidosis and fatty deposits in skeletal muscles (Vogtmann et al., 1975; Burrows and Tyrl, 2013). This concern precipitated substantive efforts to genetically reduce the erucic acid content in different rapeseed-mustard crops. Canadian rapeseed breeders were first to develop cultivars with low erucic acid (<2%) in *B. napus* and *B. rapa* (Downey and Craig, 1964). Combining low erucic acid trait with low glucosinolate content (<30 μmol/g defatted meal) in the meal subsequently aided the development of canola quality rapeseed cultivars. Similar success was achieved in *B. juncea* with the discovery of low erucic acid (<2%) genotypes (Zem 1 and Zem 2) in Australia (Kirk and Oram, 1981). Only a few varieties and hybrids of canola mustard are currently under cultivation. However, linkage drag continues to be an issue as low erucic acid donors (Zem 1 and Zem 2) and their derivatives are late maturing, small seeded, and possess low seed oil content (Banga et al., 2015). Traditional breeding methods have not been able to eliminate limitations of low oil in the presently available low erucic acid mustard varieties. Transferring the low erucic acid trait into agronomically superior genotypes is difficult, since the trait is double recessive and fatty acid composition is defined by the genotype of the embryo. Each cycle of backcrossing requires one generation of selfing after every backcross. Plants carrying low erucic acid alleles in the self-progenies are selected on the basis of half-seed analysis for the next cycle of recurrent backcrossing to the recurrent parent. A countless number of half seeds must be analyzed in each generation and this procedure is beyond the analytical capacity of most laboratories in mustard-growing countries. Although many rapid screening techniques (e.g., Near-infrared spectroscopy) are now available (Kaur et al., 2016; Sen et al., 2018), these require a large sample size and are not suitable for the analysis of single seeds. The development of functional markers with the ability to differentiate between homozygous and heterozygous states of plants is considered important. It will obviate the wet chemistry analysis for fatty acid profiling and the necessity of selfing after every generation of backcrossing.

The genes involved in the biosynthesis of very long-chain fatty acids (VLCFAs) are well known. Long-chain fatty acids are synthesized *de novo* with two enzyme systems, acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS) complex. ACCase catalyzes the formation of malonyl-acyl carrier protein (ACP) (Harwood, 1996), whereas FAS complex catalyzes the elongation of malonyl-ACP into palmitoyl-ACP with seven rounds of four sequential reactions. Elongation of palmitoyl-ACP results in stearoyl-ACP, which is then saturated by plastidal stearoyl-ACP desaturase to produce oleoyl-ACP. The removal of ACP by acyl-ACP thioesterase (FatA/FatB) marks the termination of elongation in plastids and leads to the formation of free fatty acids such as palmitic, stearic, and oleic acid. Most of the free fatty acids are transported to endoplasmic reticulum, where oleic acid is desaturated and elongated to produce polyunsaturated fatty acid and VLCFA. Fatty acid desaturases (FAD) encoded by *FAD* genes catalyze the desaturation reaction (Kunst et al., 1992; Miquel and Browse, 1992), whereas β-ketoacyl-CoA synthase (KCS) catalyzes the first step of two rounds of four sequential chain elongation reactions (James et al., 1995). This enzyme is encoded by the fatty acid elongase 1 (*FAE1*) gene in Brassica. Mutation in the *FAE1* genes impairs the activity of KCS enzyme (Katavic et al., 2002). Orthologs of the *FAE* and *FAD* genes were also reported in different oilseed *Brassica* species (Barker et al., 2007; Wang et al., 2015; Tang et al., 2019; Xue et al., 2020). *B. juncea* (AABB) and *B. napus* (AACC) possess two homologs of *FAE1.1* and *FAE1.2* genes. These show more than 99% nucleotide identity. High homologies exist between the gene sequences of the gene from *B. juncea*, *B. rapa*, and *B. napus* (Zeng and Cheng, 2014). Four and three single nucleotide polymorphisms (SNPs) in *FAE1.1* and *FAE1.2* differentiated low erucic genotypes from high erucic acid ones in *B. juncea* (Gupta et al., 2004). A total of 26 SNPs caused changes in 13 amino acids in *B. rapa* (Yan et al., 2015). Deletion of two bases in C homolog in *B. napus* produced low erucic acid phenotype (Fourmann et al., 1998). Single nucleotide primer extension (SNuPE) assays have been used to track allelic variation in the *FAE1* gene (Gupta et al., 2004), but these are expensive and tedious to implement. Cleaved amplified polymorphic sequences were also used to characterize *FAE1.1* and *FAE1.2* alleles in *B. juncea* (Saini et al., 2016), but CAPS markers are dominant and these fail to identify recessive alleles in heterozygous state. Recently, Saini et al. (2019) have reported PCR markers based on the sequence variation in the promoter regions of the *FAE1*. These are gel-based markers. A marker-assisted breeding program requires simple but robust molecular assays that benefit from economies of scale. Kompetitive allele-specific PCR (KASPar) genotyping assay is considered as an ideal approach to hasten the breeding processes through rapid genotyping of thousands of samples at very low error rates (Chen et al., 2010; Kumpatla et al., 2012). In this study, we report the development of KASPar assay for high-throughput genotyping of erucic acid content. These allele-specific assays were developed on the basis of SNPs at positions 735 and 1,476 from A and B homologs of *FAE1.1* and *FAE1.2*. These SNPs were identified on the basis of Sanger sequencing of cloned *FAE1.1* and *FAE1.2* genes from both the high and low erucic acid accessions of *B. juncea* and the progenitor species: *B. rapa* and *B. nigra*. All these sequences have been submitted to the National Center for Biotechnology Information (NCBI) and

included in **Supplementary File 1**. The practical utility of the designed KASPar assays was established by validating these in a germplasm collection and a segregating backcross progeny varying for erucic acid contents.

MATERIALS AND METHODS

Plant Material

A total of 29 pure lines of *B. juncea* (25) and inbred lines of its genome donor species, *B. rapa* (2) and *B. nigra* (2) were selected for this study (**Table 1** and **Supplementary Table 1**). Selected genotypes varied for erucic acid content (0–52%) in the oil and represented a broad spectrum of geographic and genetic diversity. The *FAE1* gene was cloned and sequenced from four genotypes of *B. juncea*, two of *B. rapa* and one of *B. nigra*. A backcross progeny 1 (BC_1) was developed from a cross of PBR357/RLC3//PBR357 [(High × Low) × High erucic genotype(s)] through hand pollinations. It was raised under field conditions and used to validate KASPar markers in segregating generation.

Fatty Acid Profiling

Fatty acid profile was determined through gas chromatography by following standard protocols (Appelqvist, 1968). For this fatty acid methyl esters (FAMEs) were injected in a fused silica capillary CP-SIL 88 column (50 mm × 0.25 mm id) fitted on gas chromatograph (Varian CP-3800, United States). Temperatures of oven, injector, and Flame ionisation detector (FID) were maintained at 200, 230, and 250°C, respectively. The peak identity of different fatty acids was determined by analyzing the FAME mix (Supelco Incorporation, United States) as a reference under similar conditions. The relative area percentage was used for estimating the identified fatty acids.

Half-Seed Technique for Non-destructive Determination of Fatty Acids

In *Brassica*, embryo genotype determines the composition of fatty acids. This mode of inheritance allows non-destructive estimation of fatty acids at a single seed level. *Brassica* seeds possess conduplicate cotyledons, where the inner cotyledon is attached with germ, and the outer cotyledon covers the inner one. The outer cotyledon was used for ascertaining the fatty acid composition and the inner one was retained for raising into adult plant. To follow half-seed method, the seeds were placed on moistened filter paper in petriplates at 25 ± 2°C under dark conditions. The swollen seeds were taken out after 24 h and the outer cotyledon of each seed was separated for estimating the fatty acid composition as per the procedure described earlier.

Cloning and Sequencing of *FAE1.1* and *FAE1.2*

Full-length *FAE1.1* and *FAE1.2* genes were isolated and cloned separately from high and low erucic acid genotypes of *B. juncea* along with its progenitors *B. rapa* and *B. nigra*. Gene-specific primers for *FAE1.1* and *FAE1.2* were designed by using previously available sequence information (NCBI Gene Accession nos. AJ558197.1 and AJ558198.1). Primers *FAE1.1* AR (forward-TGACGTCATAGTGTTAGGCGT and reverse TTTGGCACCTTTCATCGGAC) and *FAE1.2* BR (forward-ACGAAAGAGAGCAAACATCATTT and reverse-CGACA3 ACACACTGAGCAAT) were got synthesized from IDT, Germany. Thymine/Adenine (TA) cloning kit was used for cloning the amplicons in pGEM-T Easy Vector followed by sequencing with forward and reverse primer of M13. At least five clones from each genotype were sequenced by Sanger dideoxy chain termination method on capillary electrophoresis system (ABI 3730XL, Applied Biosystems, United States). Sanger sequencing was outsourced to the Eurofins Genomics India Pvt. Ltd, India. Vector contamination was removed through VecScreen tool[1]. The consensus sequences were obtained through the Geneious software (Kearse et al., 2012).

Identification of Single Nucleotide Polymorphisms

Differentiating SNPs were identified by aligning the sequences of *FAE1.1* and *FAE1.2* genes isolated from high erucic and low erucic acid accessions of *B. juncea* and its progenitor species by sequence alignment tool of the Geneious software.

Development of Kompetitive Allele-Specific PCR Assays

Gene-specific KASPar assays were manually designed for *FAE1.1* and *FAE1.2*. SNPs at positions 735 and 1,476 for *FAE1.1* and *FAE1.2* were used for the purpose (**Table 2**). Thermodynamic properties of the designed primers were analyzed by using software Vector NTI version 11.5.3[2].

Kompetitive Allele-Specific PCR Method

Kompetitive allele-specific PCR reaction mix constituted 2 μl DNA (ng), 1.944 μl KASP reagent (LGC Genomics, Beverly, MA, United States) and primer mix of 0.056 μl. PCR was run initially for 15 min at 94°C followed by 10 cycles for 20 s at 94°C and 1 min at 65°C in touchdown mode with a decrease of 1°C in every cycle and finally 30 cycles of 20 s at 94°C and for 1 min at 55°C. The Applied Biosystems™ QuantStudio™ 12K Flex (Thermo Fisher Scientific, United States) was utilized for carrying out PCR reaction and analyzing the PCR product.

RESULTS

Variation for Erucic Acid Content

Brassica *rapa* accessions, QR 2 and TL 17, contained 0.2 and 42.6% erucic acid, respectively, while both the genotypes of *B. nigra* (UP and CN 113799) had high erucic acid. *B. juncea* genotypes from India varied between 0.8 and

[1]http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html
[2]www.invitrogen.com/VectorNTI

TABLE 1 | Diverse germplasm lines of *B. juncea*, *B. rapa*, and *B. nigra*.

Genotype	Origin	Erucic acid (%)	Genotype	Origin	Erucic acid (%)
***B. rapa* (A genome)**					
QR 2#	Indian	0.2	TL 17 #	Indian	42.6
***B. nigra* (B genome)**					
U.P#	Indian	44.5	CN 113799	–	42.0
***B. juncea* (AB genomes)**					
DRMRJ 31	Indian	51.3	ALM 115	Indian	1.8
NRCDR 02	Indian	52.3	CBJ 001	Chinese	1.8
PBR 91#	Indian	45.5	EC 597325	Sweden	0.8
PBR 210	Indian	44.2	ELM 303	Indian	1.8
PBR 357	Indian	45.5	HEERA	European	2.0
PBR 378	Indian	45.0	JLM 102	Indian	1.7
Pusa Bold	Indian	45.0	JM 06013	Australian	1.5
RH 0749	Indian	52.1	PDZ 1	Indian	1.2
RH 8812	Indian	51.8	PM 24	Indian	0.8
RL 1359	Indian	47.9	PM 30	Indian	0.8
RLM 619	Indian	45.2	RLC 1	Indian	1.4
Donskaja	European	21.0	RLC 2	Indian	1.8
–	–	–	RLC 3#	Indian	0.8

#Genotypes used for sequencing.

TABLE 2 | List of the kompetitive allele-specific PCR (KASPar) primers.

Gene	Melting temperature	Amplicon size (bp)	Primer	Sequence (5′-3′)
FAE1.1	67.9	90	KASP-A-735-FAM	GAAGGTGACCAAGTTCATGCTGATGCTCTCTATGCTCACCACAAGG
	68.2		KASP-A-735-VIC	GAAGGTCGGAGTCAACGGATTGATGCTCTCTATGCTCACCACAAGA
	58.5		KASP-A-735-COM	AGTGCCGTCGTTATAGCCATTGAT
FAE1.2	67.6	108	KASP-B-1476-FAM	GAAGGTGACCAAGTTCATGCTACGAACCTCTGACTTGGCTGAATCT
	68.2		KASP-B-1476-VIC	GAAGGTCGGAGTCAACGGATTACGAACCTCTGACTTGGCTGAATCA
	58.5		KASP-B-1476-COM	TGGGTGGGTCTAAGCAATGTCAAG

52.3% for erucic acid in the oil. Erucic acid content of introduced mustard genotypes ranged between 0.8 and 21% (**Table 1**). EC 597325 and CBJ 001, of Australian and Chinese origin, respectively, exhibited low erucic acid (<2%) in the oil. The East European genotype Donskaja had intermediate erucic acid content.

Characterization of *FAE1.1*

Sequence size of *FAE1.1* from *B. rapa* cv. TL 17 (high erucic acid) and QR 2 (low erucic acid genotype) was 1,521 bp. The gene(s) comprised a single exon without any intron in *FAE1.1* from both the high and low erucic acid (**Supplementary Figure 1**). Sequence alignment of *FAE1.1* indicated high similarity (97%). Nucleotide variations were recorded at eight positions (**Supplementary Figure 2**). These variations manifested into seven transitions and one transversions. A total of 506 amino acids were encoded by messenger RNA (mRNA) of *FAE1.1* gene. Amino acid sequence comparison revealed that amino acid changes at three positions (**Supplementary Figure 2**). Four SNPs at positions 591, 735, 968, and 1,265 caused transition types of changes and were common across the available database. KASPar primers were designed only at position 735.

Characterization of *FAE1.2*

Sequence of *B. nigra* genotype, UP with high erucic acid content, was compared with the B homolog from mustard genotype, RLC3 with low erucic acid. *FAE1.2* gene had a sequence length of 1,521 bp with one exon (**Supplementary Figure 3**). Sequence alignment revealed high homology (99%) between *FAE1.2* gene sequences isolated from *B. nigra* and *B. juncea*. Nucleotide changes were identified only at three positions. These variations were indicative of two transitions and one transversion (**Supplementary Figure 4**). Translated mRNA sequence of *FAE1.2* encoded 506 amino acids. A comparison of mRNA amino acid sequences of high/low erucic acid *Brassica* genotypes indicated changes only at single position as compared to the changes at three positions for the nucleotide sequence of the gene (**Supplementary Figure 4**). Gene sequences of *FAE1.2* deciphered in our studies were also aligned with the corresponding sequences available in public domain (NCBI: AJ 558198.1). All the three identified SNPs at positions T/C (49), C/T

TABLE 3 | Single nucleotide polymorphisms (SNPs) differentiating low/high erucic acid (*FAE1.1* and *FAE1.2*) in *Brassica* species.

		SNP Position						
		FAE1.2			*FAE1.1*			
Genotypes	**Erucic acid**	**49**	**237**	**1,476**	**591**	**735**	**968**	**1,265**
B. rapa								
TL 17	49.0	C	C	T	G	C	C	T
QR 2	0.2	C	C	T	A	T	T	C
B. nigra								
UP	44.5	T	C	T	T	C	T	T
RLC 3 (B genome)	0.8	C	T	A	T	C	T	T
B. juncea								
PBR 91	45.5	T	C	T	G	C	C	T
RLC 3	0.8	C	T	A	A	T	T	C

(237), and T/A (1,476) were consensus. SNP at position 1,476 was selected for developing KASPar primer.

Identification of Single Nucleotide Polymorphisms Differentiating High/Low Erucic Acid Genotypes in *B. juncea*

Four common SNPs distinguishing high/low erucic acid *FAE1.1* gene were observed in *B. rapa* and *B. juncea* gene sequences. Likewise, three SNPs for high/low erucic acid identified for gene *FAE1.2* in *B. nigra* were also present in *B. juncea*. These SNPs were common across the ploidy and present in all the test germplasm, irrespective of geographic and genetic origin (**Table 3** and **Supplementary Figure 5**).

Development of the Kompetitive Allele-Specific PCR Assays

The KASPar assay for *FAE1.1* was designed for SNPs at position 735 and validated over 29 accessions of *B. rapa*, *B. nigra*, *and B. juncea* (**Table 4**). Genotypes with low erucic acid formed a cluster distinct from the genotypes with high erucic acid (**Figure 1A**). Amplification was not observed for *B. nigra* genotypes, indicating the A homolog specificity for the assay developed. KASPar assay specific to the gene *FAE1.2* was designed for SNP at position 1,476. This KASPar assay was also validated on 29 genotypes and formed separate clusters for high vs. low erucic acid genotypes (**Table 4** and **Figure 1B**). *B. rapa* genotypes showed no amplification, confirming the specificity of the assay for B homolog.

Validation of Kompetitive Allele-Specific PCR Assays on BC_1 Seeds

The KASPar assays were authenticated on BC_1 seeds, segregating for varied erucic acid content. The seed fatty acid composition was determined by using half-seed method. The outer cotyledon of swollen seeds was harvested for fatty acid analysis, whereas corresponding inner cotyledon with germ portion was seeded to develop into the adult plant. DNA was isolated from first true leaves of plants with known fatty acid composition. The KASPar assays were used for specific amplification of *FAE1.1* and *FAE1.2*.

TABLE 4 | Validation of the KASPar assays for erucic acid content in diverse *Brassica* genotypes.

Genotype	Erucic acid (%)	*FAE1.1* (KASPar)	*FAE1.2* (KASPar)
B. rapa			
QR 2	0.2	e1e1	NA
TL 17	49.0	E1E1	NA
B. nigra			
U.P	44.5	NA	E2E2
CN 113799	42.0	NA	E2E2
B. juncea			
DRMRIJ 31	51.3	E1E1	E2E2
NRCDR 02	52.3	E1E1	E2E2
PBR 210	44.2	E1E1	E2E2
PBR 357	45.5	E1E1	E2E2
PBR 378	45.0	E1E1	E2E2
PBR 91	45.5	E1E1	E2E2
Pusa Bold	45.0	E1E1	E2E2
RH 0749	52.1	E1E1	E2E2
RH 8812	51.8	E1E1	E2E2
RL 1359	47.9	E1E1	E2E2
RLM 619	45.2	E1E1	E2E2
Donskaja	21.0	e1e1	E2E2
ALM 115	1.8	e1e1	e2e2
CBJ 001	1.8	e1e1	e2e2
ELM 303	1.8	e1e1	e2e2
EC 597325	0.8	e1e1	e2e2
HEERA	2.0	e1e1	e2e2
JLM 102	1.7	e1e1	e2e2
JM06013	1.5	e1e1	e2e2
PDZ 1	1.2	e1e1	e2e2
PM 24	0.8	e1e1	e2e2
PM 30	0.8	e1e1	e2e2
RLC 1	1.4	e1e1	e2e2
RLC 2	1.8	e1e1	e2e2
RLC 3	0.8	e1e1	e2e2

KASP clusters showed equivalence with the segregation expected for a BC_1 progeny (**Figures 1C,D**). Four genotypic classes were inferred from the analysis of 75 BC_1 plants (**Table 5**). These

TABLE 5 | Genotyping for *FAE1.1* and *FAE1.2* genes and their association with erucic acid in *B. juncea*.

BC_1 (No. of samples)	*FAE1.1*	*FAE1.2*	*Range*	Erucic acid (Mean ± SE)
25	E1E1	E2E2	42–53	50.35 ± 0.41
14	E1E1	E2e2	38–41	39.37 ± 0.45
16	E1e1	E2E2	31–37	35.36 ± 0.35
20	E1e1	E2e2	25–30	27.29 ± 0.41
$\chi 2_{(1:1:1:1\ pvalue)}$				0.287

E1/e1, high erucic/low erucic acid allele for FAE1.1 gene; E2/e2, High erucic/low erucic acid allele for FAE1.2 gene.

were E1E1E2E2 (25): E1E1E2e2 (14): E1e1E2E2 (16): E1e1E2e2 (20). Plants with genotype E1E1E2E2 possessed average erucic acid level of 50.35% and plants heterozygous at both the loci (E1e1E2e2) had average erucic acid content of 27.29%. The remaining two genotypic classes, heterozygotes for either of two alleles, E1E1E2e2 and E1e1E2E2 showed erucic acid contents of 39.37 and 35.36%, respectively. E1 allele contributed more than that of the E2 allele. The gene action with an additive effect was indicated for the erucic acid phenotype. Heterozygotes, E1e1E2e2 showed an intermediate phenotype.

DISCUSSION

Genetic modification of oil quality is now a major breeding objective in mustard. However, most breeding groups are following traditional crop-breeding methods with little application of marker-assisted breeding in India. This is in spite of the availability of many high-density linkage maps and molecular marker systems in the species (Pradhan et al., 2003; Ramchiary et al., 2007; Panjabi et al., 2008; Yadava et al., 2012). KCS enzyme encoded by the *FAE1* is involved in one of the step for the synthesis of erucic acid from oleic acid. The mutation in coding region of the *FAE1* gene impairs its activity and results in drastic reduction of erucic acid (Katavic et al., 2002). Molecular markers, SNuPE (Gupta et al., 2004), and CAPS (Saini et al., 2016) for tracking the allelic variation for the *FAE1* gene, have been developed in *B. juncea*. Recently, a PCR-based marker system has been developed, which exploits the variation in the promoter region of the *FAE1* (Saini et al., 2019). None of these systems allow high throughput and cost-effective genotyping. It is, thus, important to develop breeder friendly and robust marker system.

We independently isolated *FAE1.1* and *FAE1.2* from four genotypes of *B. juncea*, two of *B. rapa*, and one of *B. nigra*. Cloning and gene sequencing allowed us to identify one copy each of *FAE1.1* in *B. rapa* and *FAE1.2* in *B. nigra*. Both the genes (*FAE1.1* and *FAE1.2*) were also detected in allotetraploid *B. juncea* as it has been reported earlier (Das et al., 2002; Gupta et al., 2004; Yan et al., 2015). Comparative sequence analysis has also indicated that the coding sequences (CDSs) from the *FAE1* are highly conserved with high similarity across the tested *Brassica* species. A broader genetic conservation of the gene has also been reported in the related *Brassicaceae* species, *Sinapis alba* (Zeng and Cheng, 2014). The predicted amino acid sequences of *Sinapis alba* have shared high homology with cultivated *Brassica* and *Arabidopsis*. Our studies have also emphasized the role of point/missense mutations in the coding region to inactivate KCS. Differences in the regulatory region or CDS due to SNPs, Insertion/deletions (InDels), and insertion of transposable elements cause dysfunction in the *FAE1* (Roscoe et al., 2001; Katavic et al., 2002; Chiron et al., 2015). In contrast, a deletion of four base pairs in the *FAE1* CDS from *B. napus* has produced a frameshift mutation and led to the formation of non-functional polypeptide with short length (Wu et al., 2008). Serine at 282 positions in the *FAE1* is considered essential for elongase activity in *B. napus* (Katavic et al., 2002). We have also observed transition and transversion types of nucleotide variations at eight positions in *FAE1.1* gene sequences of *B. rapa*. Of these, seven positions have been identified as transition changes and three of which cause amino acid alterations to inactivate KCS. These transition types of alterations have also been reported to cause low erucic acid in *B. rapa* (Wang et al., 2010). A total of 26 SNPs have caused changes in 13 amino acids and consequently loss of activity of KCS has also been reported in *B. rapa* (Yan et al., 2015). Whereas for *B. juncea*, total six nucleotide variations were detected in *FAE1.1* gene and all were of transition types. Comparatively, less nucleotide variations in *FAE1.1* of allotetraploid *B. juncea* were identified as compared to its diploid progenitor, *B. rapa*. Of these variations, four SNPs are common across the ploidy. These SNPs have also been reported earlier by Gupta et al. (2004) in *B. juncea*. For *FAE1.2*, nucleotide variations have occurred at the three positions both in diploid, *B. nigra* and allotetraploid, *B. juncea*. However, only one of these SNPs caused amino acid alteration that led to the loss of function of KCS. SNPs resulting from transition changes in *FAE1.2* gene have also been documented in *B. juncea* (Gupta et al., 2004). Low sequence diversity was recorded for B homologs of the *FAE1* genes in diploid and allotetraploid crop *Brassica* species. SNPs and InDels in gene or genome are being exploited for marker-assisted breeding (Garcés-Claver et al., 2007).

In this study, we planned to develop high throughput and breeder friendly markers capable of discriminating homozygous and heterozygous individuals in the segregating progenies. Effectiveness of the KASPar assays for high-throughput genotyping has already been documented in many crops including *Brassica* (Chhetri et al., 2017; Steele et al., 2018; Chen et al., 2021; Fu et al., 2021). The KASPar assays reported for *FAE1.1* and *FAE1.2* genes in the current communication were validated on large number of diverse genotypes and showed complete associations with erucic acid. High erucic acid *B. juncea* accessions are homozygous dominant for both the genes in contrast to homozygous recessive condition for low erucic acid genotypes as reported earlier (Gupta et al., 2004). Donskaja with intermediate erucic acid level is homozygous recessive for *FAE1.1* and dominant for *FAE1.2*. This observation has explained the intermediate erucic acid levels reported earlier in East European *B. juncea* (Kirk and Hurlstone, 1983). The ability

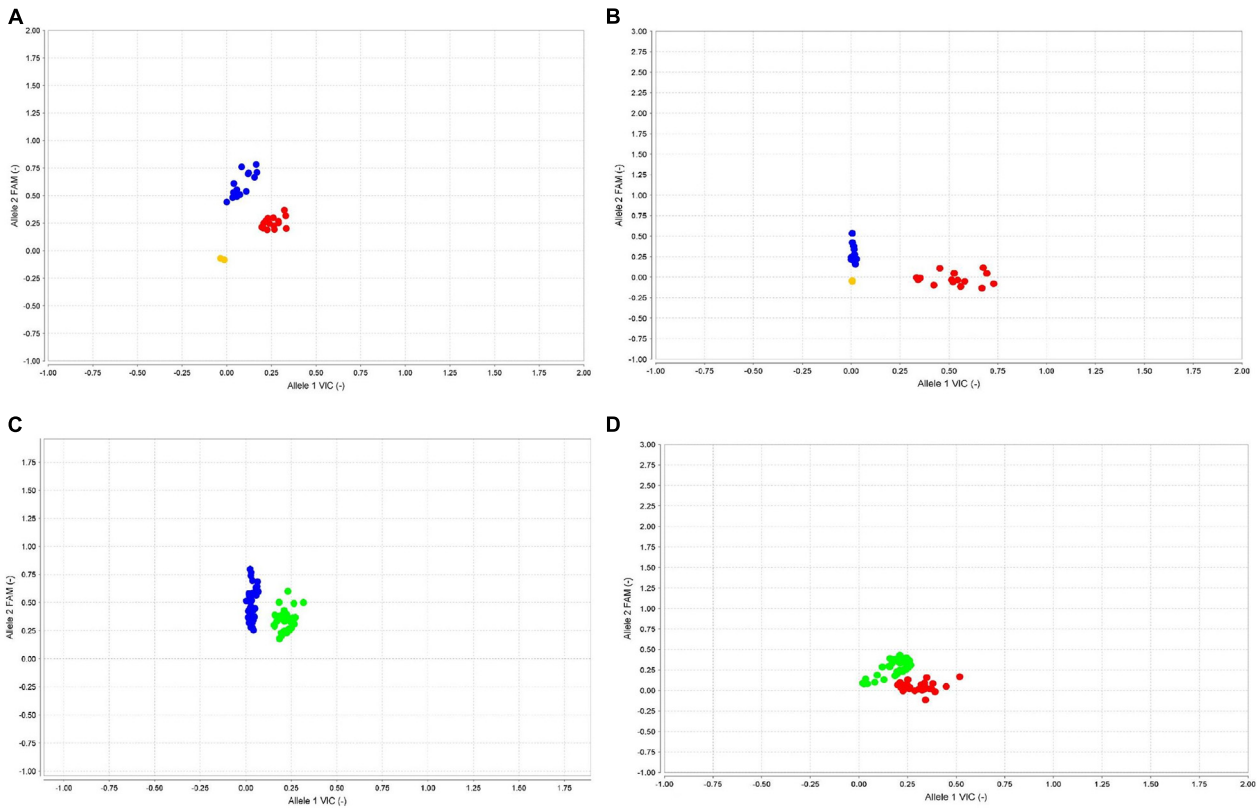


FIGURE 1 | Kompetitive allele-specific PCR (KASPar) genotyping in germplasm lines and backcross population. **(A)** KASPar genotyping of *FAE1.1* gene on pure/inbred lines: genotypes homozygous (*FAE1.1*) for high and low erucic acid clustered to the FAM side (*Y* axis: blue dots) and VIC side (*X* axis: red dots), respectively. **(B)** KASPar genotyping of *FAE1.2* gene on pure/inbred lines: genotypes homozygous for high and low erucic acid (*FAE1.2*) clustered to the VIC side (*X* axis: red dots) and FAM side (*Y* axis: blue dots), respectively. **(C)** KASPar genotyping for *FAE1.1* gene on BC_1 population. Homozygous genotypes clustered to FAM (*Y* axis: blue dots) whereas heterozygotes clusterd as green dots. **(D)** KASPar genotypes for *FAE1.2* on BC_1 population. Homozygotes clustered to VIC side (*X* axis: red dots) whereas heterozygotes clustered as green dots.

of our present marker system to differentiate between high vs. low erucic acid genotypes from the diverse geographic origins (Australia, India, China, East Europe, and Sweden) indicates its wider applicability. This could be largely due to involvement of Zem 1 and Zem 2 as donors for present low erucic acid cultivars of *B. juncea* (Potts et al., 1999; Bhat et al., 2002). These newly developed KASPar assays are codominant and could discriminate between homozygous and heterozygous states in the BC_1 progeny investigated. The segregation pattern of these assays is in confirmatory with digenic additive inheritance for erucic acid (Bhat et al., 2002; Gupta et al., 2004). The ability of our marker assays to differentiate homozygous/heterozygous state of alleles in inbred/pure lines and in segregating progenies will expedite the process of marker-assisted selection for low erucic acid in *B. juncea*.

CONCLUSION

The KASPar assays were developed on the basis of SNP variations in the coding regions of the *FAE1* genes. These assays are highly genome specific and can discriminate between heterozygous, double recessive, or double dominant states of a genotype. These can be coupled with a background selection to expedite the recovery of the recurrent parent and reduce linkage drag associated with the trait donor parent. Per sample costs are also low and a large number of samples can be processed in 1 day. Developed KASPar assays are single step, high throughput, robust, codominant, and cost-effective than gel-based markers and can be efficiently used for marker-assisted selection of low erucic acid.

DATA AVAILABILITY STATEMENT

The data set included in this manuscript is appended as **Supplementary Material**.

AUTHOR CONTRIBUTIONS

GK and SB conceived and coordinated the study. GK, SK, KK, KG, and MS developed gene-specific primers and cloned the

genes. KSG, GK, GDK, and JK conducted genotyping. GK and KSG looked after the field experiments and the phenotypic data collection. PC, MB, and GK designed the KASPar assays. KSG and GK wrote the manuscript and SB edited it. All authors read, commented, and approved the final manuscript.

FUNDING

This study was financially supported by the Indian Council of Agricultural Research under the project "Consortia for Research Platform on Molecular breeding for improvement of tolerance to biotic and abiotic stresses, yield, and quality traits in crops mustard" awarded to GK.

ACKNOWLEDGMENTS

This manuscript is based on the thesis submitted by Karanjot Singh Gill to the Punjab Agricultural University as a partial fulfilment of the requirements for the Master of Science in Plant Breeding and Genetics (https://krishikosh.egranth.ac.in/handle/1/5810039286).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.738805/full#supplementary-material

Supplementary File 1 | Nucleotide sequences of *FAE1.1* (*B. rapa and B. juncea*) and *FAE1.2* (*B. nigra* and *B. juncea*) genes in the Fast Adaptive Shrinkage Threshold Algorithm (FASTA) format.

Supplementary Figure 1 | Structure of *FAE1.1* gene of *B. rapa* genotypes **(A)** TL17 (high erucic acid) and **(B)** QR 2 (low erucic acid).

Supplementary Figure 2 | Alignment of nucleotides and amino acids derived from gene sequences of *FAE1.1* genes from high (TL17) and low erucic acid (QR 2) genotypes of *B. rapa*.

Supplementary Figure 3 | Structure of *FAE1.2* gene. **(A)** *B. nigra* (UP: high erucic acid; **B)**. *B. juncea* (RLC 3: low erucic acid).

Supplementary Figure 4 | Alignment of nucleotides and amino acids derived from gene sequences of *FAE1.2* genes from high (UP) and low erucic acid (RLC 3) genotypes of *B. nigra and B. juncea*, respectively.

Supplementary Figure 5 | Alignment of nucleotides of *FAE1.1* and *FAE1.2* genes of *B. juncea* (PBR91: high erucic acid and RLC 3: low erucic acid).

Supplementary Table 1 | Source of diverse germplasm lines of *B. juncea*, *B. rapa*, and *B. nigra*.

REFERENCES

Appelqvist, L. A. (1968). Lipids in Cruciferae: III. Fatty acid composition of diploid and tetraploid seeds of Brassica campestris and Sinapisalba grown under two climatic extremes. *Plant Physiol.* 21, 615–625. doi: 10.1111/j.1399-3054.1968.tb07286.x

Banga, S. K., Kumar, P., Bhajan, R., Singh, D., and Banga, S. S. (2015). "Genetics and Breeding" in *Brassica Oilseeds: Breeding and Management*. eds A. Kumar, S. S. Banga, P. D. Meena, and P. R. Kumar (Wallingford: CABI), 11–41. doi: 10.1079/9781780644837.0011

Barker, G. C., Larson, T. R., Graham, I. A., Lynn, J. R., and King, G. J. (2007). Novel Insights into Seed Fatty Acid Synthesis and Modification Pathways from Genetic Diversity and Quantitative Trait Loci Analysis of the Brassica C Genome. *Plant Physiol.* 144, 1827–1842. doi: 10.1104/pp.107.096172

Bhat, M. A., Gupta, M. L., Banga, S. K., Raheja, R. K., and Banga, S. S. (2002). Erucic acid heredity in Brassica juncea. *Plant Breed.* 121, 456–458. doi: 10.1046/j.1439-0523.2002.731805.x

Burrows, G. E., and Tyrl, R. J. (2013). *Toxic Plants of North America*, 2nd Edn. New Jersey: Wiley. doi: 10.1002/9781118413425

Chen, W., Mingus, J., Mammadov, J., Backlund, J. E., Greene, T., Thompson, S., et al. (2010). "KASPar: a simple and cost-effective system for SNP genotyping" in *Proceedings of Plant and Animal Genomes XVII Conference*. (US). 194.

Chen, Z., Tang, D., Ni, J., Li, P., Wang, L., Zhou, J., et al. (2021). Development of genic KASP SNP markers from RNA-Seq data for map-based cloning and marker-assisted selection in maize. *BMC Plant Biol.* 21:157. doi: 10.1186/s12870-021-02932-8

Chhetri, M., Bariana, H., Wong, D., Sohail, Y., Hayden, M., and Bansal, U. (2017). Development of robust molecular markers for marker-assisted selection of leaf rust resistance gene Lr23 in common and durum wheat breeding programs. *Mol. Breed.* 37:21. doi: 10.1007/s11032-017-0628-6

Chiron, H., Wilmer, J., Lucas, M. O., Nesi, N., Delseny, M., Devic, M., et al. (2015). Regulation of Fatty acid elongation1 expression in embryonic and vascular tissues of Brassica napus. *Plant Mol. Biol.* 88, 65–83. doi: 10.1007/s11103-015-0309-y

Das, S., Roscoe, T. J., Delseny, M., Srivastava, P. S., and Lakshmikumaran, M. (2002). Cloning and molecular characterization of the Fatty Acid Elongase 1 (FAE1) gene from high and low erucic acid lines of Brassica campestris and Brassica oleracea. *Plant Sci.* 162, 245–250. doi: 10.1016/S0168-9452(01)00556-8

Downey, R. K., and Craig, B. M. (1964). Genetic control of fatty acid biosynthesis in rapeseed (Brassica napus L). *J. Am. Oil Chem. Soc.* 41, 475–478. doi: 10.1007/BF02670026

Fourmann, M., Barret, P., Renard, M., Pelletier, G., Delourme, R., and Brunel, D. (1998). The two genes homologous to Arabidopsis FAE1 co-segregate with the two loci governing erucic acid content in Brassica napus. *Theor. Appl. Genet.* 96, 852–858. doi: 10.1007/s001220050812

Fu, Y., Mason, A. S., Zhang, Y., and Yu, H. (2021). Identification and Development of KASP Markers for Novel Mutant BnFAD2 Alleles Associated With Elevated Oleic Acid in Brassica napus. *Front. Plant Sci.* 12:142. doi: 10.3389/fpls.2021.715633

Garcés-Claver, A., Fellman, S. M., Gil-Ortega, R., Jahn, M., and ArnedoAndrés, M. S. (2007). Identification validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in Capsicum spp. *Theor. Appl. Genet.* 115, 907–916. doi: 10.1007/s00122-007-0617-y

Gupta, V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y. S., Pental, D., and Pradhan, A. K. (2004). Molecular tagging of erucic acid trait in oilseed mustard (Brassica juncea) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. *Theor. Appl. Genet.* 108, 743–749. doi: 10.1007/s00122-003-1481-z

Harwood, J. L. (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochim. Biophys. Acta Lipids Lipid Metab.* 1301, 7–56. doi: 10.1016/0005-2760(95)00242-1

James, D. W., Lim, E., Keller, J., Plooy, I., Ralston, E., and Dooner, H. K. (1995). Directed tagging of the Arabidopsis fatty acid elongation 1 (FAE1) gene with the maize transposon activator. *Plant Cell* 7, 309–319. doi: 10.1105/tpc.7.3.309

Katavic, V., Mietkiewska, E., Barton, D. L., Giblin, E. M., Reed, D. W., and Taylor, D. C. (2002). Restoring enzyme activity in nonfunctional low erucic acid Brassica napusfatty acid elongase 1 by a single amino acid substitution.Eur. *J. Biochem.* 269, 5625–5631. doi: 10.1046/j.1432-1033.2002.03270.x

Kaur, B., Sangha, M. K., and Kaur, G. (2016). Calibration of NIRS for the estimation of fatty acids in Brassica juncea. *J. Am. Oil Chem. Soc.* 93, 673–680. doi: 10.1007/s11746-016-2802-0

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software

platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kirk, J. T. O., and Hurlstone, C. G. (1983). Variation and inheritance of erucic acid content in Brassica juncea. *Z Pflanzenzucht* 90, 331–338.

Kirk, J. T. O., and Oram, R. N. (1981). Isolation of erucic acid-free lines of Brassica juncea: indian mustard now a potential oilseed crop in Australia. *J. Aust. Inst. Agric. Sci.* 47, 51–52.

Kumpatla, S. P., Buyyarapu, R., Abdurakhmonov, I. Y., and Mammadov, J. A. (2012). "Genomics-assisted plant breeding in the 21st century: technological advances and progress," in *Plant Breed*. ed. I. Y. Abdurakhmonov (InTech). 131–183. doi: 10.5772/37458

Kunst, L., Taylor, D. C., and Underhill, E. W. (1992). Fatty acid elongation in developing seeds of Arabidopsis thaliana. *Plant Physiol. Biochem.* 30, 425–434.

Miquel, M., and Browse, J. (1992). Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis.Biochemical and genetic characterization of a plant oleoyl-phosphatidyl choline desaturase. *J. Biol. Chem.* 267, 1502–1509. doi: 10.1016/S0021-9258(18)45974-1

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

Potts, D. A., Rakow, G. W., and Males, D. R. (1999). "Canola-quality Brassica juncea, a new oilseed crop for the Canadian prairies" in *New Horizons for an old crop. Proceedings of 10th International Rapeseed Congress*. (Australia).

Pradhan, A., Gupta, V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y., and Pental, D. (2003). A high-density linkage map in Brassica juncea (Indian mustard) using AFLP and RFLP markers. *Theor. Appl. Genet.* 106, 607–614. doi: 10.1007/s00122-002-1083-1

Ramchiary, N., Padmaja, K. L., Sharma, S., Gupta, V., Sodhi, Y. S., Mukhopadhyay, A., et al. (2007). Mapping of yield influencing QTL in Brassica juncea: implications for breeding of a major oilseed crop of dryland areas. *Theor. Appl. Genet.* 2115, 807–817. doi: 10.1007/s00122-007-0610-5

Roscoe, T. J., Lessireb, R., Puyaubertb, J., Renardc, M., and Delsenya, M. (2001). Mutations in the fatty acid elongation 1 gene are associated with a loss of L-ketoacyl-CoA synthase activity in low erucic acid rapeseed. *FEBS Lett.* 492, 107–111. doi: 10.1016/S0014-5793(01)02243-8

Saini, N., Koramutla, M. K., Singh, N., Singh, S., Singh, R., Yadav, S., et al. (2019). Promoter polymorphism in FAE1.1 and FAE1.2 genes associated with erucic acid content in Brassica juncea. *Mol. Breed.* 39:75. doi: 10.1007/s11032-019-0971-x

Saini, N., Singh, N., Kumar, A., Vihan, N., Yadav, S., Vasudev, S., et al. (2016). Development and validation of functional CAPS markers for the FAE genes in Brassica juncea and their use in marker-assisted selection. *Breed. Sci.* 66, 831–837. doi: 10.1270/jsbbs.16132

Sen, R., Sharma, S., Kaur, G., and Banga, S. S. (2018). Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. *J. Sci. Food Agr.* 98, 4050–4057. doi: 10.1002/jsfa.8919

Steele, K. A., Quinton-Tulloch, M. J., Amgai, R. B., Dhakal, R., Khatiwada, S. P., Vyas, D., et al. (2018). Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice. *Mol. Breed.* 38:38. doi: 10.1007/s11032-018-0777-2

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

Vogtmann, H., Christian, R., Hardin, R. T., and Clandinin, D. R. (1975). The effects of high and low erucic acid rapeseed oils in diets for rats. *Int. J. Vitam. Nutr. Res.* 45, 221–229.

Wang, N., Shi, L., Tian, F., Ning, H., Wu, X., Long, Y., et al. (2010). Assessment of FAE1 polymorphisms in three Brassica species using EcoTILLING and their association with differences in seed erucic acid contents. *BMC Plant Biol.* 10:137. doi: 10.1186/1471-2229-10-137

Wang, X., Long, Y., Yin, Y., Zhang, C., Gan, L., Liu, L., et al. (2015). New insights into the genetic networks affecting seed fatty acid concentrations in Brassica napus. *BMC Plant Biol.* 15:91. doi: 10.1186/s12870-015-0475-8

Wu, G., Wu, Y., Xiao, L., Li, X., and Lu, C. (2008). Zero erucic acid trait of rapeseed (Brassica napusL.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. *Theor. Appl. Genet.* 116, 491–499. doi: 10.1007/s00122-007-0685-z

Xue, Y., Jiang, J., Yang, X., Jiang, H., Du, Y., Liu, X., et al. (2020). Genome-wide mining and comparative analysis of fatty acid elongase gene family in Brassica napus and its progenitors. *Gene* 747:144674. doi: 10.1016/j.gene.2020.144674

Yadava, S. K., Arumugam, N., Mukhopadhyay, A., Sodhi, Y. S., Gupta, V., Pental, D., et al. (2012). QTL mapping of yield-associated traits in Brassica juncea: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. *Theor. Appl. Genet.* 125, 1553–1564. doi: 10.1007/s00122-012-1934-3

Yan, G., Li, D., Cai, M., Gao, G., Chen, B., Xu, K., et al. (2015). Characterization of FAE1 in the zero erucic acid germplasm of Brassica rapaL. *Breed. Sci.* 65, 257–264. doi: 10.1270/jsbbs.65.257

Zeng, F., and Cheng, B. (2014). Transposable Element Insertion and Epigenetic Modification Cause the Multiallelic Variation in the Expression of FAE1 in Sinapisalba. *Plant Cell* 26, 2648–2659. doi: 10.1105/tpc.114.126631

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 in Plant Science

ORIGINAL RESEARCH
published: 16 September 2021
doi: 10.3389/fpls.2021.721631

Genetic Analysis of Heterosis for Yield Influencing Traits in *Brassica juncea* Using a Doubled Haploid Population and Its Backcross Progenies

Aakanksha[1†], *Satish Kumar Yadava*[2†], *Bal Govind Yadav*[1], *Vibha Gupta*[2], *Arundhati Mukhopadhyay*[2], *Deepak Pental*[2] *and Akshay K. Pradhan*[1,2*]

[1] Department of Genetics, University of Delhi South Campus, New Delhi, India, [2] Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi, India

OPEN ACCESS

Edited by:
Jianjun Chen,
University of Florida, United States

Reviewed by:
Berisso Demo Kebede,
University of Alberta, Canada
Reeta Bhatia,
Indian Agricultural Statistics Research Institute, India

***Correspondence:**
Akshay K. Pradhan
pradhancgmcp@gmail.com

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

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

Received: 07 June 2021
Accepted: 16 August 2021
Published: 16 September 2021

Citation:
Aakanksha, Yadava SK, Yadav BG, Gupta V, Mukhopadhyay A, Pental D and Pradhan AK (2021) Genetic Analysis of Heterosis for Yield Influencing Traits in Brassica juncea Using a Doubled Haploid Population and Its Backcross Progenies. Front. Plant Sci. 12:721631. doi: 10.3389/fpls.2021.721631

The exploitation of heterosis through hybrid breeding is one of the major breeding objectives for productivity increase in crop plants. This research analyzes the genetic basis of heterosis in *Brassica juncea* by using a doubled haploid (DH) mapping population derived from F_1 between two heterotic inbred parents, one belonging to the Indian and the other belonging to the east European gene pool, and their two corresponding sets of backcross hybrids. An Illumina Infinium Brassica 90K SNP array-based genetic map was used to identify yield influencing quantitative trait loci (QTL) related to plant architecture, flowering, and silique- and seed-related traits using five different data sets from multiple trials, allowing the estimation of additive and dominance effects, as well as digenic epistatic interactions. In total, 695 additive QTL were detected for the 14 traits in the three trials using five data sets, with overdominance observed to be the predominant type of effect in determining the expression of heterotic QTL. The results indicated that the design in the present study was efficient for identifying common QTL across multiple trials and populations, which constitute a valuable resource for marker-assisted selection and further research. In addition, a total of 637 epistatic loci were identified, and it was concluded that epistasis among loci without detectable main effects plays an important role in controlling heterosis in yield of *B. juncea*.

Keywords: ***Brassica juncea*, heterosis, QTL mapping, yield influencing traits, dominance, epistasis**

INTRODUCTION

Heterosis is defined as the superior performance of F_1 hybrids relative to the homozygous parents (East, 1908; Shull, 1908). While the practical application of heterosis in plant breeding is quite successful in many crops through the development of hybrid varieties, the molecular mechanism of the phenomenon is not well understood. The development of molecular quantitative genetics has facilitated the study of the genetic basis of heterosis in crops (Paterson et al., 1988; Stuber, 1992). Numerous studies have been carried out in different crops, including maize, which is an outcrossing crop, and rice and wheat, which are self-pollinated. Three main hypotheses exist

to explain the genetic basis of heterosis—dominance, overdominance, and epistasis (Crow, 1999; Goodnight, 1999; Lippman and Zamir, 2007). The dominance hypothesis supposes that deleterious recessive alleles of one of the parents are complemented in the F_1 hybrid by the dominant alleles of the other parent. Results from several quantitative genetic experiments in crops like rice and maize favor the dominance hypothesis (Xiao et al., 1995; Cockerham and Zeng, 1996; Swanson-Wagner et al., 2009). The overdominance hypothesis states that the heterozygous combination of the alleles at a locus is superior to either of the two possible homozygous combinations. Overdominance has been seen as the primary genetic basis of heterosis in different crops like maize (Stuber et al., 1992), rice (Li et al., 2001; Luo et al., 2001; Mei et al., 2005; Zhu et al., 2016), tomato (Semel et al., 2006; Krieger et al., 2010), and cotton (Ma et al., 2019). Epistasis is defined as the interactions between alleles of different loci and has been shown as the major determinant of heterosis in a few studies (Yu et al., 1997; Hua et al., 2002, 2003; Luo et al., 2009; Tang et al., 2010; Zhou et al., 2012).

In *Brassica napus*, Radoev et al. (2008) used doubled haploids (DH) and the corresponding backcross lines with their midparent heterosis data to map heterotic quantitative trait loci (QTL) and identified a large number of epistatic interactions for seed yield and three yield-component traits. They concluded that epistasis, together with all levels of dominance from partial to overdominance, was responsible for the expression of heterosis in rapeseed. Using the same experimental design in two DH populations and two corresponding sets of backcross hybrids, Basunanda et al. (2010) detected a number of QTL hotspots, which consisted of epistatic loci as well as main-effect QTL associated with seedling biomass and yield-related traits. They found that epistatic interactions played an important role in the expression of heterosis in oilseed rape and postulated that these QTL hotspots might harbor genes involved in the regulation of heterosis for different traits, with an overall significant influence on heterosis for seed yield. Prominent clustering of QTL for yield and yield-related traits have also been reported in genomes of many other crops and model plants, such as rice (You et al., 2006), maize (Frascaroli et al., 2007), wheat (Peng et al., 2003), mustard (Ramchiary et al., 2007; Yadava et al., 2012), and Arabidopsis (Fu et al., 2009), indicating that QTL for yield-related traits might have pleiotropic effects. Studies on various crop species have consistently postulated that the genetic mechanism of heterosis is complex without any single explanation for its expression (Radoev et al., 2008; Liang et al., 2015; Shang et al., 2016; Li et al., 2018; Ma et al., 2019; Liu et al., 2020).

Different mapping populations with varied genetic structures have previously been used to study heterosis in crops (Schnable and Springer, 2013). F_2 and BC populations can be used to study heterosis; however, their temporary nature in the form of single plants as the segregating units restricts their use in experiments with replications (Yu et al., 1997; Li et al., 2012). To circumvent these issues, new population types in the form of "immortalized F2" (IF2) populations and double backcrosses of DH or Recombinant inbred lines (RILs) have been constructed in several studies (Xiao et al., 1995; Hua et al., 2002, 2003; Mei et al., 2005; Shi et al., 2011). Several studies on heterosis using a backcross design have been carried out in different crops, such as rice (Xiao et al., 1995; Li et al., 2001; Luo et al., 2009), maize (Frascaroli et al., 2007; Larièpe et al., 2012), and rapeseed (Radoev et al., 2008; Basunanda et al., 2010).

Oilseed mustard, *Brassica juncea* (AABB), an important oilseed crop of the Indian subcontinent is grown under low-moisture regimes during the winter-growing season. The major breeding objectives in *B. juncea* target yield enhancement, oil quality, resistance against pathogens, and tolerance for abiotic stresses (Pradhan and Pental, 2011). Extensive phenotypic and DNA marker-based studies on oilseed types of *B. juncea* have confirmed the existence of two major distinct gene pools—-Indian and East European (Pradhan et al., 1993; Srivastava et al., 2001). Hybrids between the Indian and east European types are heterotic for yield (Pradhan et al., 1993). In *B. juncea*, the first ever commercial hybrid DMH-1 released in India, developed from a cross between an Indian gene pool line Pusa bold and an east European gene pool line EH-2, shows high heterosis with about 30% higher yield than the best check varieties (Sodhi et al., 2006). The molecular basis of heterosis in *B. juncea* has not been investigated so far, and there remain several unresolved basic questions on the genetics of this highly complex trait. The principal objective of this study was, therefore, the genetic analysis of heterosis by identification, localization, and estimation of the effects of QTL for seed yield and related component traits, along with an exhaustive assessment of the contributions of different genetic effects, e.g., dominance, overdominance, and epistasis, to the expression of heterosis in *B. juncea*.

MATERIALS AND METHODS

Plant Materials

A DH mapping population was derived from a cross between an archetypical Indian gene pool line of mustard—Varuna, and EH-2, an early flowering mutant of Heera (the canola-quality exotic eastern European line). The two lines are highly divergent and good heterotic combiners belonging to two distinct gene pools—Indian and east European. Briefly, crosses were made between Varuna and EH-2, using Varuna as the female parent. F_1 plants were grown in controlled environmental chambers for the isolation of microspores to obtain DH plants through *in vitro* colchicine treatment of microspores and embryogenesis (Mukhopadhyay et al., 2007). Second or third leaf from 3-to-4-week-old individual putative doubled haploid plants was taken for analyzing its ploidy level, following Arumuganathan and Earle (1991), and the diploidized plants were transplanted in the field.

The DH population (hereinafter referred to as "VEH") consisted of 164 lines developed from the F_1 of a cross between Varuna and EH-2. Two backcross (BC) populations were developed from crosses between the DH lines with Varuna and with EH-2 by using the DH lines as the female parent, and Varuna or EH-2 as the male parent, hereinafter referred to as "BC-V" and "BC-E," respectively. Each of these two BC populations consisted

of 164 BCF_1 hybrids corresponding to the 164 DH lines of the VEH population.

Field Experiments and Data Analysis

The DH lines, the backcross populations, the two parental lines and the F_1 of the Varuna/EH-2 cross were grown in fields located in Delhi during the crop seasons of 2014 to 2015, 2015 to 2016, and 2016 to 2017 (**Supplementary Table 1**), following a randomized complete block design with three replications in each of the three trials. Phenotyping for yield-influencing traits was performed with five competitive plants from each replication, and the mean of the 15 observations was used as the trait value. The traits were broadly classified as plant architectural traits and days to flowering, silique-related traits, and seed-related traits as described in literature (Cai et al., 2016; Zhao et al., 2019, 2021). A total of 14 characters defining plant architecture, days to flowering, silique-related traits, and seed-related traits were evaluated in the three mapping populations: (1) plant height (PH); (2) days to 50% flowering (DF); (3) number of primary branches (PBR); (4) number of secondary branches (SBR); (5) main shoot length (MSL); (6) single plant yield (SPY); (7) silique on the main shoot (SQMS); (8) silique length (SQL); (9) seeds in a silique (SSQ); (10) silique density (SQD); (11) number of siliques on a plant (SQPL); (12) thousand seed weight (TSW); (13) oil content (OIL); and (14) protein content (PRO).

The levels of midparent heterosis were estimated for the F_1 hybrid of the parents "Varuna" and "EH-2," referred to as F_1-heterosis. For testing the significance of heterosis values, *t*-tests were applied. The data were statistically analyzed using Plabstat (Utz, 2004), and the trait ANOVAs were performed using the phenotyping module of iMAS (Integrated Marker Assisted Selection) version 2.1 (http://www.icrisat.org/gt-bt/imas.htm). The adjusted mean (Best Linear Unbiased Prediction, BLUP) values across the three environments were calculated with a mixed linear model using the R package metan (Olivoto and Lúcio, 2020), accounting for effects of environment, replication, genotype, and genotype by the environment.

Genetic Mapping

Young leaves were collected from the parents, F_1 and DH lines, and stored at −80°C. Total genomic DNA was isolated from the leaves, following the CTAB method (Rogers and Bendich, 1994). Polymorphic markers available in the lab were used for genotyping the VEH DH mapping population and were used to develop the framework linkage map. Two types of agarose gel-based markers —Intron Polymorphism (IP) (Panjabi et al., 2008) and simple sequence repeat (SSR) (Dhaka et al., 2017)—and a third type of markers—SNP—identified by PCR-based KASP assays, which were previously developed using the transcriptome data for two *B. juncea* varieties —Varuna and Heera (Paritosh et al., 2014)—were used in this study. These IP, SSR, and KASP-SNP markers served as anchor markers, which were used to identify the linkage groups and their positions on the VEH linkage map corresponding to their previously known positions on the 10 A sub-genome and 8 B sub-genome linkage groups of the *B. juncea* linkage map, following Paritosh et al. (2014).

For the development of an SNP-based high-density genetic map of *B. juncea*, Illumina Infinium Brassica 90K SNP array was used as it could generate large amount of marker data in a single experiment with reasonably good distribution across the genome and high-information content, including physical positions. The availability of a large number of markers genotyped on the VEH population entailed genetic mapping in the R statistical computing environment using the ASMap package (Taylor and Butler, 2017; R Core Team, 2020). Genotypes of the backcross hybrids constituting the two BC populations were inferred from the genotypes of the VEH lines.

Correlation Analysis

Phenotypic data were organized into five data sets: DH (VEH), two backcrosses (BC-V and BC-E), and two corresponding midparent heterosis data sets (MPH-V and MPH-E). Correlation coefficients between trait values in five data sets, BLUPs, and those between genome heterozygosity and trait performance were estimated in the R statistical environment (R Core Team, 2020). Mean of the trait values of three replications of each of the data set in individual trials and BLUPs for the combined analysis were used to calculate correlation coefficients between five data sets.

To calculate correlation coefficients between genome heterozygosity and trait performance in the BC populations, genome heterozygosity for the two BC populations was assessed from the genome ratios of the 164 doubled haploid lines. Genome heterozygosity for the BC-V (VEH × Varuna) population was calculated as the percentage of genome, which originated from EH-2, and that for BC-E (VEH × EH-2) population, as the percentage of a genome that originated from Varuna.

QTL Analysis

The five data sets (VEH, BC-V, BC-E, MPH-V, and MPH-E) were used separately for QTL mapping. The midparent heterosis data sets were estimated as the deviation of the backcross hybrids (*BC-V* and *BC-E*) from the midparent value (*MPV*), derived as the mean of the corresponding doubled haploid line (*DH*) and the male parent (Varuna or EH-2). The following equations were used for midparent heterosis estimation, following Radoev et al. (2008):

$$MPV_i = (DH_i + P1 \text{ or } P2)/2$$
$$MPH_i = BC_i - MPV_i$$

where DH is the VEH doubled haploid line, P1 is Varuna, P2 is EH-2, and BC is a backcross hybrid.

Quantitative trait loci, which contribute to heterosis, are presumably those detected with the *MPH*-data sets.

For each data set, the trait averaged values of three replications in each of the three environments were used for QTL analysis, while the combined analysis was based on BLUPs for each trait across the three environments. QTL mapping was carried out using CIM (composite interval mapping) in Windows QTL Cartographer 2.5 (Wang et al., 2012). For CIM, the standard model (Model 6) was used with forward regression, a window size of 10 cM and five background control markers. The genome was scanned for putative QTL with main effects

at a walking speed of 1 cM. The experiment-wise error rate was determined by performing 1,000 permutations to obtain the empirical thresholds (Churchill and Doerge, 1994). We adopted LOD values >3.0 for identifying "significant" QTL. To avoid missing minor QTL, a lower LOD value corresponding to $p \leq 0.20$ was adopted for the presence of "suggestive" QTL (Shi et al., 2011). For each trait, multiple QTL peaks detected from a single trial within 10 cM of each other were regarded as a single QTL. The QTL were classified as major when their R^2 (phenotypic variation explained) value was more than 10%. The graphical representations of QTL on linkage groups were drawn using MapChart 2.32 software (Voorrips, 2002).

The gene actions in different data sets varied according to the type of the mapping population and the input data. The three types of data sets (DH, BC, and MPH) described above provided different genetic effect information. QTL obtained in the DH population explained the additive effect "a." The dominance effect (d) was explained by the QTL obtained in two midparent heterosis data sets, MPH-V, and MPH-E, which revealed the heterotic loci. QTL obtained in the backcross populations explained a combination of additive and dominance effects – (a + d)/2 and (a–d)/2 if the donor or the recurrent parent carries a dominant-increasing allele, respectively (Radoev et al., 2008). QTL detected simultaneously in different data sets allowed an assessment of the degree of dominance by calculating d/a ratio. The mode of inheritance of identified QTL was defined as additive ($|d/a|<0.2$), partially dominant ($0.2\leq|d/a|<0.8$), dominant ($0.8\leq|d/a|<1.2$), and overdominant ($|d/a|\geq1.2$) (Stuber et al., 1987).

To dissect the epistasis (QTL × QTL) involved in the expression of the traits in the VEH, BC-V, BC-E, MPH-V, and MPH-E data sets, software IciMapping 4.0 (Meng et al., 2015) was used. A threshold of LOD ≥ 2.5 (PIN = 0.001) was used for declaring the presence of main-effect QTL (M-QTL), and LOD ≥ 5.0. (PIN = 0.0001) was used for declaring the presence of epistatic QTL (E-QTL) (Shang et al., 2016). In the case of epistasis, the estimated effects in VEH population are the additive × additive genetic interactions, while the effects calculated in the BC-V, BC-E, MPH-V, and MPH-E data sets are complex mixtures of all possible epistatic interactions: additive × additive (aa), additive × dominance (ad), and dominance × dominance (dd) interactions (Radoev et al., 2008).

Candidate Gene Analysis

Sequences of the SNP markers flanking QTL regions were mapped to the *B. juncea* reference genome (Paritosh et al., 2021; https://www.ncbi.nlm.nih.gov/assembly/GCA_015484525.1), available on NCBI database using BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to project the QTL on the physical map and to find marker positions. Genes that were physically located within the QTL regions were annotated using the TAIR GO term enrichment tool (https://www.arabidopsis.org/tools/go_term_enrichment.jsp), which sends the data to Panther classification system (Ashburner et al., 2000; Thomas et al., 2003; Mi et al., 2010, 2021).

RESULTS

Construction of the VEH Linkage Map

A total of 513 polymorphic markers (IP, SSR, and KASP-SNP) were used as anchor markers in the linkage analysis. A further set of 10,007 high throughput SNPs were identified out of the 77,969 markers available on the ABC 90K Illumina chip. Of these, markers with missing data (> 30%), significant segregation distortion, and monomorphic markers were removed. This screen left us with 7,274 robust markers along with the 513 anchor markers that were used for constructing the final map. Out of the 7,787 markers that were selected for genotyping the VEH population, 7,609 markers (7,101 SNPs and 508 anchor markers) showed consistent ordering on the 18 linkage groups (**Table 1**). The total length of the map was 2,627.9 cM with an average number of 422 markers per linkage group (**Supplementary Table 2**).

Trait Performance of Parents, F_1, VEH, and Two Backcross Populations

The performance of VEH population, two backcross populations (BC-V and BC-E), parental lines (Varuna and EH-2), and their F_1 hybrid, along with the levels of midparent heterosis (MPH) observed in the F_1, BC-V, and BC-E populations for the 14 traits are summarized in **Supplementary Table 3**. The two parents contrasted for most of the traits, making them ideal candidates for heterosis breeding and related experiments. In the F_1, significant midparent heterosis was observed for PH, PBR, SBR, SPY, SQD, SQPL, and OIL.

The BC-V population showed significantly high MPH for SBR, SPY, and SQPL, while, in the BC-E population, significantly high MPH was recorded for PBR, SBR, SPY, SQD, and SQPL. Significant genetic variation was observed in most of the traits in the VEH, BC-V, and BC-E populations (**Supplementary Table 4**). DF, MSL, and PH showed high-genetic variance and heritability in all three trials of the populations. High heritabilities were also observed for OIL, SQD, SQL, and TSW even with low genetic variances in all trials in the VEH as well as in BC-V and BC-E populations. Siliques on the main shoot and siliques on a plant showed high genetic variance but lower heritability in all trials of all populations. The trait PBR showed low genetic variance but high heritability in two of the three trials. PRO showed low genetic variance in all three populations with high heritability in VEH but lower in both BC populations. The traits SBR and SPY showed high-genetic variance in two trials of VEH population but reduced genetic variances in all the trials of the two BC populations. SSQ showed high-genetic variance in VEH population, lower in the BC populations while showing moderate to high heritability in all the populations and trials.

Correlations Among Traits

Correlation coefficients among the 14 traits were estimated using the phenotypic data of the three environments and BLUPs. Both positive and negative correlations among the traits were observed in all the three populations in the three trials (**Supplementary Table 5**). In VEH, PH showed significant

TABLE 1 | Characteristic features of the linkage groups in the VEH map, constructed using IP, SSR, KASP-SNP, and Chip-based SNP markers.

Chromosome	Anchor markers	Chip SNPs	Number of markers	Total length (cM)	Average spacing (cM)	Marker interval	Marker density
A1	33	376	409	110.7	0.3	76	3.7
A2	40	394	434	135.3	0.3	82	3.2
A3	39	388	427	139.1	0.3	93	3.1
A4	27	402	429	104.5	0.2	62	4.1
A5	28	347	375	116.3	0.3	73	3.2
A6	29	181	210	142.7	0.7	72	1.5
A7	21	375	396	138.8	0.4	61	2.9
A8	15	202	217	99.0	0.5	71	2.2
A9	43	568	611	171.6	0.3	104	3.6
A10	12	171	183	122.8	0.7	53	1.5
A Genome	287	3,404	3,691	1280.8	–	747	–
Mean	28.7	340.4	369.1	128.1	0.4	74.7	2.9
B1	24	788	812	112.0	0.1	54	7.3
B2	23	758	781	174.1	0.2	110	4.5
B3	38	816	854	206.7	0.2	133	4.1
B4	28	80	108	162.2	1.5	54	0.7
B5	24	238	262	157.5	0.6	82	1.7
B6	16	176	192	133.3	0.7	62	1.4
B7	36	432	468	154.4	0.3	91	3.0
B8	32	409	441	247.0	0.6	103	1.8
B Genome	221	3,697	3,918	1347.1	–	689	–
Mean	27.6	462.1	489.8	168.4	0.5	86.1	3.1
Total (A+B Genome)	508	7,101	7,609	2627.9	–	1,436	–
Mean	28.2	394.5	422.7	146.0	0.3	79.8	3.0

positive correlations with DF, PBR, SQD, and SQMS. The traits PBR and SBR showed a significant negative correlation with MSL and TSW. Single plant yield showed a significant positive correlation with PBR, SBR, SQMS, SQPL, and TSW. Similar correlations were observed in BC-V and BC-E populations.

Correlations Between DH Lines, Backcross Performance, and Midparent Heterosis

In all the traits, a significantly high positive correlation was observed between the performances of VEH and the two BC populations in all three trials (**Supplementary Table 6**). Phenotypic performance of VEH for most traits was negatively correlated with MPH performances. Performances of BC-V and BC-E populations showed a significantly high positive correlation with the corresponding MPH performances.

Relationships Between Traits and Genome Heterozygosity

Relationships between molecular marker heterozygosity and phenotypic performance of the 14 traits in the BC hybrids was also evaluated by regressing the midparent heterosis value and the trait value of each BC hybrid on its percentage of genome heterozygosity. Heterozygosity in the backcross populations showed a mean of 49 and 48% in the BC-V and BC-E populations, respectively. For most traits, genome heterozygosity was found to be positively correlated with the performance of the two BC populations and MPH (**Table 2**). The correlation coefficients were, however, mostly low in the four data sets (BC-V, MPH-V, BC-E, and MPH-E), indicating that the overall genome heterozygosity alone had little effect on the trait expression.

Main-Effect QTL

A total of 695 main-effect QTL were detected for the 14 traits in the three trials using the five data sets (**Supplementary Table 7**). Of these, 176 QTL were detected in VEH, 188 in BC-V, 164 in BC-E, 92 in MPH-V, and 75 in the MPH-E data set. Approximately 75% of these 695 QTL were significant and were identified with LOD values >3, while the remaining proportion consisted of suggestive QTL with LOD values ranging between 2.5 and 3.

A highest number of QTL were detected on the linkage groups B3 (86), A7 (84), and A3 (77) while the lowest number of QTL was detected on B6 (5). The total number of major QTL explaining a phenotypic variance of more than 10% was 175, with the highest number of these mapping to A7 (**Supplementary Table 7**). A summary of heterotic QTL and their effects is shown in **Table 3**. Overdominance was observed to be the predominant type of effect in the heterotic QTL.

The combined analysis using BLUPs identified a total of 281 QTL from the five data sets, of which approximately 68% were identified as significant with LOD values >3. The largest number of QTL was detected in VEH population with a total of 72 QTL. Similar to the number of QTL identified with MPH-E data set in the single environments, MPH-E detected the least number of 31 QTL in the combined analysis also (**Supplementary Table 8**).

TABLE 2 | Coefficients of correlation between whole-genome heterozygosity and trait performance in the BC and MPH data sets.

Trait	Year	BC-V	MPH-V	BC-E	MPH-E
Plant architectural traits and days to flowering					
PH	2014-15	0.276**	0.278**	−0.049	0.186*
	2015-16	0.285**	0.332**	−0.002	0.171*
	2016-17	0.258**	0.192*	−0.083	0.163
DF	2014-15	0.118	0.105	−0.109	−0.013
	2015-16	0.173*	0.204**	0.089	−0.091
	2016-17	0.071	0.082	−0.024	0.02
MSL	2014-15	−0.042	0.03	−0.005	−0.114
	2015-16	−0.033	0.041	−0.037	−0.131
	2016-17	−0.1	−0.041	0.002	−0.057
PBR	2014-15	0.171*	0.107	−0.009	0.07
	2015-16	0.101	0.031	0.084	0.09
	2016-17	0.186*	0.199*	0.057	0.167
SBR	2014-15	0.300**	0.310**	0.087	0.083
	2015-16	0.116	0.099	0.069	0.063
	2016-17	0.081	0.026	0.206*	0.222**
Silique related traits					
SPY	2014-15	0.365**	0.377**	0.241**	0.186*
	2015-16	0.082	0.095	0.104	0.07
	2016-17	0.141	0.153	0.102	0.166*
SQD	2014-15	0.164*	0.062	−0.062	0.170*
	2015-16	0.217**	0.02	0.002	0.268**
	2016-17	0.207*	0.15	0.076	0.171*
SQL	2014-15	−0.297**	0.055	0.343**	0.077
	2015-16	−0.377**	−0.051	0.379**	0.083
	2016-17	−0.283**	0.004	0.230**	−0.005
SQMS	2014-15	0.159*	0.101	−0.08	0.068
	2015-16	0.189*	0.09	−0.044	0.078
	2016-17	0.174*	0.189*	0.122	0.145
SQP	2014-15	0.397**	0.323**	0.126	0.133
	2015-16	0.261**	0.143	0.089	0.135
	2016-17	0.222**	0.170*	0.046	0.15
SSQ	2014-15	0.276**	0.145	−0.034	0.016
	2015-16	0.231**	0.107	−0.005	0.072
	2016-17	0.212*	−0.024	−0.083	0.133
Seed related traits					
OIL	2014-15	0.270**	0.300**	0.084	0.164*
	2015-16	0.069	0.141	0.092	0.114
	2016-17	0.192*	0.228**	0.065	0.13
PRO	2014-15	−0.115	−0.201**	−0.168*	−0.126
	2015-16	−0.156*	−0.276**	−0.237**	−0.174*
	2016-17	−0.082	−0.239**	−0.215*	−0.113
TSW	2014-15	−0.349**	−0.108	0.225**	−0.123
	2015-16	−0.418**	−0.248**	0.223**	−0.127
	2016-17	−0.350**	−0.142	0.187*	−0.031

*Significant at 5%, **Significant at 1%.

TABLE 3 | Effects of heterosis QTL identified for 14 yield traits in two backcross populations in the three trials.

		BC-V				BC-E			
	Trait	**A**	**PD**	**D**	**OD**	**A**	**PD**	**D**	**OD**
Plant architectural traits and days to flowering	PH	9	2	3	13	6	2	1	10
	DF	14	2	4	11	9	1	2	5
	MSL	8	1	1	3	8	0	2	9
	PBR	6	0	0	9	5	0	0	5
	SBR	8	0	0	9	6	0	0	5
Silique related traits	SPY	3	0	0	7	6	0	0	6
	SQD	8	0	1	8	8	0	1	11
	SQL	14	1	3	9	6	1	4	5
	SQMS	7	0	0	4	2	0	0	5
	SQPL	6	0	1	7	3	0	1	7
	SSQ	7	0	2	3	5	2	1	2
Seed related traits	OIL	8	3	1	7	7	0	2	9
	PRO	8	0	1	5	11	1	0	4
	TSW	6	0	0	13	13	1	0	4
	Total	112	9	17	108	95	8	14	87

A, additive effect QTL; PD, partial dominance effect QTL; D, dominance effect QTL; and OD, over dominance effect QTL.

Plant Architectural Traits and Days to Flowering

Plant Height

Twenty-two QTL with additive effects were detected in the VEH population in the three trials. The QTL on LG A07 were consistently identified in all three trials and showed a large positive additive effect with substantial proportions of phenotypic variance explained (PVE). For all other QTL detected for PH, the additive effects were mostly negative, indicating that the alleles for taller plants were contributed by the parent EH-2 (**Supplementary Table 7**).

A total of 21 and 17 QTL were identified in the BC-V and BC-E populations of backcross hybrids, respectively, in the three trials. All of the six common QTL detected in the BC-V, BC-E, and VEH populations showed considerably larger effects in the backcross populations compared with the additive effects observed in VEH at these loci, indicating non-additive gene action. The two MPH data sets detected 14 QTL, of which three QTL in MPH-V and two in MPH-E were also found in other data sets. The effects in the two MPH data sets were mostly positive, suggesting that these QTL in heterozygous conditions would increase PH with respect to the midparent value, consistent with the positive heterosis observed for this trait. For QTL detected simultaneously in the different data sets, degree of dominance could be calculated, which revealed that the trait is controlled by a combination of mostly overdominant loci with a few partially dominant loci.

Days to Flowering

Seventeen QTL were mapped in the VEH population in the three trials, 14 of which showed negative additive effects, indicating that Varuna contributed the alleles for an earlier flowering time. Consistent with the highly significant correlations between DF and PH, major QTL on the LG A07 were identified as common QTL for both the traits, with Varuna contributing the positive alleles for taller plants and delayed flowering. The two backcross populations identified a total of 35 QTL in the three trials. Fifteen QTL in BC-V and 11 QTL in BC-E showed negative effects, indicating that the alleles for delayed flowering in the hybrid combination were provided by EH-2. The QTL analysis for the two MPH data sets identified 12 QTL with PVEs ranging from 5 to 12.5%. Two of the major QTL for DF detected in MPH-V showed dominance effects with opposite signs. Thus, the counterbalancing dominance effects along with the additive effects explain the absence of significant heterosis in DF.

Main Shoot Length

Twelve QTL were detected for MSL in the VEH population, which explained PVEs ranging from 4.3 to 15.3% in the three trials. In seven cases, the effects were positive, implying that Varuna contributed the positive alleles for MSL at these loci. Twenty-four QTL were mapped with the two backcross data sets, and the QTL on LGs A04, A07, and B03 were also identified in the doubled haploid (VEH) population. Ten QTL were detected with the two MPH data sets with an equal number of loci showing positive and negative dominance effects, thus, resulting in the low heterosis exhibited by the trait.

Number of Primary Branches (PBR)

A total of 38 QTL were identified, including 16 major QTL in the five data sets. Fifteen QTL were detected in the VEH population for which both the parents, Varuna, and EH-2 contributed positive alleles. Of the 19 QTL mapped in the two backcross populations, 14 showed negative effects, indicating that the increasing alleles were provided by EH-2. Fourteen QTL were identified using the MPH-V and MPH-E data sets, and the effects were mostly positive, consistent with the significant heterosis

observed for PBR. The QTL on LG A07 were consistently detected as major QTL with high PVEs in BC-V, BC-E, VEH, and MPH-E data sets, wherein Varuna increased the number of primary branches in both homozygous and heterozygous conditions at this locus.

Number of Secondary Branches (SBR)

An equal number of seven loci each, with positive and negative effects among the total of 14 QTL, were detected for SBR, indicating the dispersal of increasing alleles in both Varuna and EH-2. Of the 17 QTL mapped in the BC-V and BC-E populations, 13 showed negative effects, indicating that EH-2 contributed the trait-increasing alleles for number of secondary branches at these loci. Eleven QTL were identified in the two MPH data sets, seven of which showed overdominance as they were not identified in the backcross and DH data sets.

Silique-Related Traits

Single Plant Yield

Ten QTL were mapped with the VEH data set, with both Varuna and EH-2 contributing the positive alleles at an equal number of five loci each (**Supplementary Table 7**). Three of these were identified as major QTL, explaining more than 10% of the PVE, and were located on the LGs A04, B02, and B04. The two backcross data sets detected 16 QTL, of which five were detected as major QTL. At four of these five major QTL identified in the BC-V and BC-E populations, Varuna provided the increasing dominant allele. Fifteen dominant QTL were detected with the MPH-V and MPH-E data sets, and 13 of these showed positive dominance effects, implying that the heterozygotes exhibited higher single plant yields congruent with the highly significant amount of heterosis observed for SPY. The effects of these dominant loci were also appreciably greater than the effects observed in the DH (VEH) population, indicating their overdominant nature.

Silique Density

Twelve QTL affecting SQD were detected in the VEH population in the three trials, with the higher parent (EH-2) donating the increasing allele at eight of these loci. The two backcross populations detected 29 QTL in the three trials, which were distributed as 13 loci in BC-V and 15 loci in BC-E, along with a common locus for the trait. The effects were mostly negative in BC-V and positive in BC-E, indicating that EH-2 and Varuna provided the trait-enhancing alleles as donor parents in BC-V and BC-E, respectively. Fourteen QTL were mapped with the MPH-V and MPH-E data sets, and nine of these showed positive dominance effects, indicating that these loci would increase SQD in heterozygous state. This observation is in consonance with the positive heterosis shown by the trait. Furthermore, seven of the 14 QTL detected with the two MPH data sets showed overdominance as no corresponding QTL were identified with additive effects in other data sets.

Silique Length

Fifteen QTL with additive effects were identified in the VEH population in the three trials. Except the QTL on the LG A01, which received the trait-increasing allele from EH-2, all the remaining 14 QTL inherited the trait-increasing allele for SQL from the higher parent, Varuna. Most of the 28 QTL identified in the two backcross populations also showed positive effects, indicating the contribution of the increasing-dominant allele by Varuna as the recurrent parent in BC-V and as the donor parent in the BC-E hybrids. Nine QTL were detected in the MPH-V and MPH-E data sets, of which three over-dominant QTL showed negative dominance effects. The low heterosis observed for SQL was thus attributed to QTL, with mostly additive effects and a few other over-dominant loci with opposite signs.

Siliques on the Main Shoot

Six QTL were detected in the VEH population in the three trials, of which four QTL showed negative effects, indicating the contribution of positive alleles by EH-2 toward increasing siliques on the main shoot. Fourteen QTL were detected in the two backcross populations, of which 10 QTL showed negative effects, implying the role of EH-2 in donating the trait-enhancing alleles as the donor parent in BC-V and as the recurrent parent in BC-E population. Eleven QTL were identified in the MPH-V and MPH-E data sets, all of which were considered as overdominant QTL. Two major QTL with large effects were detected in the LGs A05 and B02, with opposite signs. Based on the QTL effects, SQMS appeared to be influenced largely by additive effects in combination with a few overdominant loci, which showed counterbalancing effects.

Siliques on a Plant

All the seven QTL identified in the VEH population invariably showed negative additive effects, indicating that the higher parent EH-2 contributed the alleles at these loci for an increase in the number of siliques on a plant. In the two backcross populations, 10 QTL were detected in the BC-V population compared with only four in the BC-E population. All the QTL in the BC-V population along with the two QTL identified in the BC-E population exhibited negative effects, implying that EH-2 provided the increasing dominant alleles at these loci as the donor parent in BC-V and as the recurrent parent in BC-E. Most importantly, a total of 13 QTL identified with the two MPH data sets showed positive dominant effects along with QTL on LG B02, which showed a negative dominance effect. The QTL identified in the MPH-E data set and mapping to LG A09 exhibited the largest effect, which was about six times the effect observed for the smallest effect QTL in the VEH population. The significantly high heterosis observed for SQPL can be explained with a large number of loci exhibiting overdominance.

Seeds in a Silique

Seven QTL with PVEs ranging from 6.3 to 22.2% were identified in the VEH population in the three trials. Six of these, including the major QTL, showed negative effects, indicating that these QTL inherited the alleles for the increased number of seeds in a silique from the higher parent, EH-2. The two backcross populations detected a total of 18 QTL, including six major QTL with PVEs >10%. Similar to the results obtained with the

VEH population, a majority of 15 QTL in the BC-V and BC-E displayed negative effects, implying that EH-2 contributed to the increasing dominant alleles in both populations. The QTL mapping with the two MPH data sets resulted in the detection of six minor QTL, all of which showed negative dominance effects. The trait of SSQ is most probably controlled by loci with additive effects as the number of QTL with dominance effects was less as compared with the loci with additive effects, an observation consistent with the insignificant heterosis observed for SSQ.

Seed-Related Traits

Oil Content

Eleven QTL with significant additive effects were mapped in the VEH population in the three trials, with an approximately equal number of six and five positive alleles for increased oil content being contributed by Varuna and EH-2, respectively (**Supplementary Table 7**). The QTL on the linkage groups on A08 and B07 were consistently detected in all the three trials. The genomic regions harboring these QTL have previously been shown to contain the erucic acid genes (Paritosh et al., 2014; Rout et al., 2018). All the five QTL in these regions showed positive effects, indicating that Varuna contributed to the increasing alleles. For the remaining six QTL, EH-2 donated the positive alleles. A total of 24 QTL were detected in the two backcross populations in the three trials, two-thirds of which showed negative effects, indicating the contribution of EH-2 in increasing oil content in the backcross hybrids. The QTL analysis with the two MPH data sets revealed the existence of 17 QTL, showing dominance effects. Of these 17 loci, a majority of 14 QTL showed positive dominance effects, indicating that heterozygosity at these QTL would increase oil content, which was congruent with the significant levels of heterosis for this trait. For the QTL on the LGs A08 and B07, both the additive and dominance effects were positive, meaning that, at these loci, the Varuna allele for increasing oil content was dominant.

Protein Content

Of the 14 QTL identified for PRO in the VEH population in the three trials, 10 showed negative effects, indicating that EH-2 contributed alleles for higher protein content in the doubled haploid lines. Twenty-three QTL were identified in the two backcross populations in the three trials. While the effects were mostly positive in BC-V, the effects in BC-E were negative, indicating that the recurrent genotype had a major influence on the protein content of the backcross hybrids. A total of six QTL were mapped using the two MPH data sets, four of which showed negative dominance effects in agreement with the negative but insignificant heterosis observed for the trait in two of the three trials. The QTL on LG A03 showed both the positive additive and dominance effects, indicating that the Varuna allele-increasing protein content was dominant. The mode of genetic action thus appears to be mostly due to additive loci along with a few dominant loci with opposite signs.

Thousand Seed Weight

Fourteen QTL were mapped in the VEH population, a majority of which showed positive effects, indicating that the parent Varuna contributed the favorable alleles. Twenty-seven QTL were detected in the BC-V and BC-E hybrids and explained PVEs ranging from 4.8 to 19.5%. The genetic effects observed in the BC-V population were invariably positive, while, in the BC-E population, 13 of the 17 QTL showed positive effects. The prevalence of the positive effects in the two backcross populations can be ascribed to the trait-increasing alleles provided by the higher parent, Varuna. A comparatively greater number of 10 QTL were identified using the MPH-V data set compared with only two with the MPH-E data set. In MPH-V, all the QTL, excluding one, displayed negative dominance effects, indicating that these QTL in the heterozygous state would reduce seed weight. Conversely, both the QTL detected with MPH-E data set exhibited positive dominance effects, connoting a positive influence of heterozygosity on seed weight at these loci. However, the negative dominance effects can be presumed to result in the observed negative heterosis for seed weight. The mode of action governing TSW in *B. juncea* thus appears to be a combination of additive effects and QTL in the dominance-overdominance range.

Environmentally Stable QTLs

Quantitative trait loci, which were detected within two or all of the three environments for the five data sets, and also in the combined analysis based on BLUPs, were considered as "environmentally stable QTL." Based on this criterion, a total of 88 environmentally stable QTL were identified for all the 14 traits in the five data sets (**Supplementary Table 8**). The largest number of 11 environmentally stable QTL was detected for PH, while only two QTL for SQMS were identified concurrently in the combined analysis and in at least two of the three environments. Consistent with the QTL analysis in the three individual environments, the genetic intervals on A07 were identified as the environmentally stable QTL for all the traits, excluding SQMS and PRO.

Epistatic QTL

A total of 637 epistatic QTL (E-QTL) were detected in the five data sets for the 14 traits (**Supplementary Table 9**). Out of these, 3 interactions were found to be between two main-effect QTL (Type I interaction), 102 interactions between main-effect QTL and QTL without any significant main effect (Type II interaction), and 532 interactions were between two QTL without any significant main effect (Type III interaction) (**Table 4**).

In the VEH population, a total of 262 epistatic interactions involving 505 loci were identified with PVEs in the range of 0.4–8.7%. The average PVE of the epistatic interactions was 2.8%. Of the 14 traits analyzed for epistasis in the VEH population, SSQ was characterized by the highest number of 41 interactions, while SQPL detected only 4 interactions. SSQ also showed the highest number of 8 Type II and 33 Type III interactions. A total of 49 main-effect QTL were observed to contribute to the Type I and Type II interactions. Only a single Type I interaction was identified in the VEH population, which was identified for PH. All of the digenic interactions observed for SPY exhibited negative effects, denoting recombinant allele combinations increased SPY.

TABLE 4 | Type of epistatic interactions detected in the five data sets in 14 traits.

Trait	Population/Data set	Type 1	Type 2	Type 3	Total E-QTL
Plant architectural traits and days to flowering					
PH	BC-V	0	5	18	23
	BC-E	0	5	19	24
	VEH	1	8	23	32
	MPH-V	0	0	5	5
	MPH-E	0	1	2	3
DF	BC-V	0	2	6	8
	BC-E	0	3	12	15
	VEH	0	6	24	30
	MPH-V	0	1	7	8
	MPH-E	0	0	6	6
MSL	BC-V	0	0	7	7
	BC-E	0	1	4	5
	VEH	0	2	22	24
	MPH-V	0	0	2	2
	MPH-E	0	0	0	0
PBR	BC-V	1	5	9	15
	BC-E	0	3	4	7
	VEH	0	1	9	10
	MPH-V	0	0	5	5
	MPH-E	0	0	0	0
SBR	BC-V	0	3	14	17
	BC-E	0	0	6	6
	VEH	0	5	16	21
	MPH-V	0	3	11	14
	MPH-E	0	0	1	1
Silique related traits					
SPY	BC-V	0	0	8	8
	BC-E	0	0	5	5
	VEH	0	2	4	6
	MPH-V	0	0	2	2
	MPH-E	0	0	2	2
SQD	BC-V	0	2	5	7
	BC-E	0	1	9	10
	VEH	0	2	14	16
	MPH-V	0	0	1	1
	MPH-E	0	0	2	2
SQL	BC-V	0	3	4	7
	BC-E	0	2	6	8
	VEH	0	2	10	12
	MPH-V	0	0	3	3
	MPH-E	0	0	1	1
SQMS	BC-V	0	1	17	18
	BC-E	0	0	3	3
	VEH	0	1	23	24
	MPH-V	0	1	8	9
	MPH-E	0	0	1	1
SQP	BC-V	0	3	4	7
	BC-E	0	0	10	10
	VEH	0	0	4	4
	MPH-V	0	0	3	3
	MPH-E	0	0	2	2

(Continued)

TABLE 4 | Continued

Trait	Population/Data set	Type 1	Type 2	Type 3	Total E-QTL
SSQ	BC-V	0	0	12	12
	BC-E	0	2	7	9
	VEH	0	8	33	41
	MPH-V	0	0	4	4
	MPH-E	0	0	5	5
Seed related traits					
OIL	BC-V	0	1	5	6
	BC-E	0	1	9	10
	VEH	0	3	12	15
	MPH-V	0	0	5	5
	MPH-E	0	2	1	3
PRO	BC-V	0	2	4	6
	BC-E	0	1	7	8
	VEH	0	2	9	11
	MPH-V	0	0	4	4
	MPH-E	0	0	4	4
TSW	BC-V	1	1	10	12
	BC-E	0	0	2	2
	VEH	0	5	11	16
	MPH-V	0	0	4	4
	MPH-E	0	0	1	1
Total		3	102	532	637

Totally, 275 interactions were identified in the BC-V and BC-E populations, with an average PVE of 2.4% in each of these two populations. The PVEs by the interacting loci were in the range of 0.07 to 5.7% in BC-V and 0.6 to 4.4% in the BC-E population. In the BC-V population, a total of 295 loci were involved in 2, 14, and 123 interactions belonging to Types I, II, and III, respectively. In comparison, 236 loci were involved in a total of 122 epistatic interactions of Types II and III in the BC-E population. The interacting E-QTL involved 32 and 19 QTL with main effects in the BC-V and BC-E populations, respectively. Notably, PH showed the highest number of interactions in both the backcross populations, majority of which were Type III, occurring between background loci with no significant main effects. The traits of SPY and SQMS were characterized by negative epistatic interactions, indicating that the recombination among the parental alleles would increase SPY and SQMS.

The two MPH data sets, MPH-V, and MPH-E individually detected 69 and 31 digenic interactions, respectively. The interactions identified using the MPH-V data set exhibited PVEs in the range of 0.06 to 5.6%, with an average of 2.9%, while the loci involved in interactions identified using the MPH-E data set showed PVEs, ranging from 0.09 to 5.7%, with an average of 2.9%, similar to that observed with the MPH-V data set. The traits of SBR in MPH-V and DF in the MPH-E data set displayed the largest number of 14 and 6 interactions, respectively. The traits, SPY and SQPL in the MPH-V, and SBR and SQMS in MPH-E included interactions with negative effects only, meaning the recombinant allele combinations favored increases in the phenotypes of these traits. Ninety-three percent of the interactions, which were detected using the two MPH data sets, were of Type III, occurring between background loci without significant main effects, while the remaining 7% were of Type II and involved a combination of QTL with main effect and a modifying background locus.

Active Regions

Quantitative trait loci clusters or hotspots were also identified for the 14 yield and related traits, using the five data sets of all the three trials (**Supplementary Figure 1**). QTL hotspots have been previously defined as active regions (Basunanda et al., 2010). A total of 59 such active regions were detected in this study, with a maximum of five such regions, each mapping to LGs A09 and B08. The most notable active region was found on the LG A7, which consisted of 79 QTL in a genetic interval of 0.0 cM-72.8 cM. It harbored QTL for 12 of the 14 traits, excluding SQMS and SQPL. Another dense active region was detected on LG A03, consisting of 75 QTL, located in the genetic interval from 20.3 cM to 138.8 cM and had QTL from all the 14 traits. An active region, which included 65 QTL from 13 traits, was also identified on LG B03, the only exception being QTL for SQL. It appears that these regions harbor "reference" or "housekeeping" genes that are expressed during the entire life cycle of the plant by influencing a large number of traits related to plant architecture, flowering, etc. (Basunanda et al., 2010).

Quantitative trait loci analysis performed with BLUPs also showed the presence of an active region on A7 and resulted in a more precise interval of 0.0–11.2 cM with QTL for 8 of the 14 traits. This region, consisting of QTL for days to flowering, plant architectural traits, silique- and seed-related traits was therefore, analyzed *in silico* to identify putative candidate genes. Based on the annotation of *Arabidopsis* genome, 288 genes were identified, which were subsequently analyzed using TAIR GO term enrichment tool. The analysis revealed a total of 67 significant genes ($p < 0.05$) under the GO Biological processes category, which included MYB122 (AT1G74080), bHLH160 (AT1G71200), NAC031 (AT1G76420), MYB54 (AT1G73410), and other transcription factors (Riechmann et al., 2000; Stracke et al., 2001; Vroemen et al., 2003), gene-encoding single-stranded DNA-binding protein WHY2 (AT1G71260), which is a mitochondrial member of the WHIRLY family of plant-specific proteins (Desveaux et al., 2005), WAT1 (AT1G75500), which is required for secondary wall formation in wood fibers (Ranocha et al., 2013), PIN1(AT1G73590), and PIN3 (AT1G70940), both of which act as components of the auxin efflux carrier (Zhou and Luo, 2018) (**Supplementary Table 10**).

DISCUSSION

Advantages of the Experimental Design for Genetic Analysis of Heterosis

This study provides the first report on the use of different segregating populations from the same parental combination to study the genetics of heterosis for yield in *B. juncea*. The two inbred parents, Varuna and EH-2, have been shown to be good heterotic combiners, as the F_1 hybrid from this combination (hybrid DMH-11) shows about 28–36% yield heterosis in plot-level yield trials (personal communication). The unique experimental strategy of using the DH and the two BCF_1 (BC-V and BC-E) populations was specifically designed to accurately resolve different types of gene action along with a comprehensive mapping of loci contributing to yield-influencing traits in *B. juncea*. The two backcross populations can be repeatedly developed from the DH lines, and the genotypes of the backcross hybrids can be unambiguously deduced from the marker information of the parental doubled haploid lines, thus obviating the need for additional genotyping and mapping procedures. A similar design was earlier adopted by Radoev et al. (2008) and Basunanda et al. (2010) for studying heterosis for seedling biomass and yield traits.

Zhihong et al. (2012) have also highlighted the advantages of this design, also known as the double backcross design and is used for constituting an immortal mapping population, allowing for across-environment replications, estimation of dominance effects, epistatic effects, and QTL-environment interactions, and thus remedies the drawbacks of a single backcross population. Recent studies on QTL mapping in rice, maize, cotton, and *B. napus* have shown that dominance and epistasis, especially the additive × additive interaction, play a key role in contributing to heterosis (Stuber et al., 1992; Yu et al., 1997; Tang et al., 2010). Using a conventional BC population, the additive × additive interactions cannot be detected separately because epistasis is identified as mixtures of additive and dominance effects. The double backcross population design obviates this problem by accurately and efficiently estimating additive × additive interaction effects.

The results obtained in this study revealed that the design in the present study was more efficient to map common stable QTL and several heterozygous loci across multiple populations than that could be possibly detected from a single-base population. However, the backcross populations could account for only 50% of the possible heterosis. Use of the two BC populations was also made possible to clearly detect dominant QTL from both the parents, which would otherwise have been difficult if we had used independent testers (Mei et al., 2003).

A High-Density Linkage Map in *B. juncea*

Several linkage maps have been constructed in *B. juncea* using mapping populations with various genetic structures based on a wide array of molecular markers, and a number of QTL for yield and quality traits have also been identified from these linkage maps (Ramchiary et al., 2007; Pradhan and Pental, 2011; Yadava et al., 2012; Yang et al., 2016; Dhaka et al., 2017; Rout et al., 2018; Paritosh et al., 2021). However, heterosis for yield and yield component traits has not been analyzed using any of these maps reported so far. The VEH linkage map includes a large number of SNPs, which are the most frequent type of genetic polymorphisms, providing a high density of markers near a locus of interest (Picoult-Newberg et al., 1999). The high-density map used in this study was based on SNP markers identified using ABC 90k Illumina SNP array (Mason et al., 2017). The well-characterized sequence-based markers used in the array, along with the availability of *B. juncea* genome sequence, ensure the efficient dissection and subsequent use of complex traits in Brassica breeding programs.

The Illumina 60K Brassica SNP Bead Chip array (Clarke et al., 2016) has previously been used in QTL analysis in *Brassica napus* to study flowering time, seeds per silique, silique length, seed weight, and other traits (Yang et al., 2017; Wu et al., 2018; Wang et al., 2019; Song et al., 2021) but has not yet been utilized in genetic mapping of heterotic loci. Because *B. juncea* and *B. napus* share A genome, therefore using common markers, an attempt was made to align QTL for yield-related traits detected in the present study with those detected in previous studies undertaken in *B. napus*. However, only one QTL detected for seeds per silique on linkage group A07 with PVE >10% (**Supplementary Table 7**) could be aligned with QTL (qSSA07.4) identified by Yang et al. (2017). Furthermore, the IP and SSR markers used in the present study have previously been used to construct a high-density VH map and different integrated maps using multiple populations for studying various seed and oil traits. Based on common markers, three QTL detected for thousand seed weight—one each on A3 (PVE >10%, 102.8–113 cM), A4 (60.9–75.5 cM), and A7 (48.2–64.2 cM)—were aligned with the QTL detected previously by Dhaka et al. (2017), and three oil-content QTL—one each on A7 (10.9–24.1 cM), A8 (12.6–42.2 cM), B7 (77–93.6 cM)—were aligned with QTL detected previously by Rout et al. (2018). The high-density map of *B. juncea* developed in this study will,

therefore, facilitate QTL congruency studies by evaluating QTL detected in different mapping populations.

Mapping QTL With Additive and Dominance Effects

In this study, the performance of BCF_1 hybrids of the BC-V and BC-E populations was compared with the performance of the parental DH lines (VEH) based on both the phenotypic correlations (**Supplementary Table 5**) and QTL analyses (**Supplementary Table 7**). A total of distinct 336 additive QTL were identified for 14 traits in the DH and two BC populations in the three experiments. Of these, 100 were identified only in the DH population, while 131 and 105 QTL were unique to the BC-V and BC-E populations, respectively. Thus, Varuna was the superior tester than EH-2, which probably carried dominant alleles at several loci, and thus, reduced the ability to detect QTL in the BC-E data set. Twenty-six QTL showed strong additive effects and were consistently identified among the three (VEH, BC-V, and BC-E) data sets, while only seven QTL were observed to be common between the BC-V and BC-E populations, which can be primarily attributed to the mode of prevailing gene action (**Supplementary Table 7**). For these common QTL detected in the three data sets, the direction of contribution was identical. These findings are consistent with the assumption that, in case of the additive model, the heterozygote is exactly halfway between the two homozygotes, and hence, the effect of allele substitution can be revealed in more than one TC population (Melchinger et al., 1998; Frascaroli et al., 2007, 2009).

The BLUP method of analysis (Henderson, 1974) considers fixed environmental and random genetic effects at the same time, and therefore, improves the accuracy of the BLUP values for phenotypic data obtained in different years, different locations, and of progenies of different generations (Piepho et al., 2008; Wang et al., 2016). Using combined QTL analysis based on the BLUP values across three environments, we were able to map environmentally stable QTL. Although a comparatively reduced number of QTL were detected in the combined analyses, the QTL identified using BLUPs corroborated the analyses based on single environments, and a majority of the QTL (78%) were common with the QTL obtained in the analyses of at least one environment (**Supplementary Tables 7, 8**). In the combined analysis, 51 significant QTL were detected using the BC-V data set compared with 41 significant QTL detected with BC-E data set, indicating the superiority of Varuna over EH-2 as a tester in the current study.

A number of additive QTL were detected either in the BC populations or in VEH population. For QTL detected in the DH population but not in the BC-V, BC-E or the two MPH data sets, the dominant allele carried by the recurrent parent can be assumed to be fully dominant over the donor parent, and *vice versa*. The effects at these loci are based on the difference between the additive and dominance effects, which cancel each other in case of full dominance or they are too low to be detected in case of partial dominance. Conversely, there were QTL observed only in the BC-V and BC-E hybrids (and not in the VEH and the two MPH data sets) (**Supplementary Table 7**). The genetic effects for these loci represent the sum of the additive and dominance effects, which are probably, too low to be detected separately in the other data sets (Radoev et al., 2008).

Heterosis can be estimated over better parent also when it is called better parent heterosis. In the current study, we have, however, identified QTL using the two MPH data sets (MPH-V and MPH-E) only, as midparent heterosis is one of the most adopted ways of reporting heterosis. For a deeper understanding of the genetic causes of heterosis, comparison of the hybrid with the average performance of both parental lines (and not with only the better parent) remains the most acceptable strategy because the F_1 hybrid inherits half its nuclear genome from each parent (Melchinger et al., 2007).

A comparatively smaller number of QTL for all the traits were detected in the MPH-V and MPH-E data sets with a total of 92 and 74 QTL, respectively. One of the reasons for this may be the presence of QTL with an intermediate mode of inheritance, which, lacking dominance, could not be detected in MPH data. Secondly, QTL with additive effects larger than dominance effects, that is, QTL with partial or incomplete dominance, were less likely to be identified in MPH data than in the VEH and BC-V or BC-E data (Radoev et al., 2008; Shang et al., 2016). A total of 79 QTL identified in the two MPH data sets were congruent with loci identified with the other data sets, which allowed the assessment of the degree of dominance (**Table 3**). Of the 134 dominance effects estimated in the BC-V population in the three trials, 9 showed partial dominance, 17 showed dominance, and the remaining 108 exhibited overdominance. In BC-E, out of 109 dominance effects that were estimated, 8 showed partial dominance, 14 showed dominance, while the remaining 87 showed overdominance. A number of QTL identified exclusively with MPH data sets and loci, exhibiting dominance effects higher than the lowest additive effect identified for a trait with VEH population, were suggestive of overdominance. Single-plant yield, which showed the highest level of heterosis, was characterized by a total of six overdominant QTL in the two MPH data sets. In *B. napus*, the largest number of loci exhibiting overdominance was identified for grain yield (Radoev et al., 2008), and seed yield and seed number per plant (Shi et al., 2011) showing the highest level of midparent heterosis of these traits.

In a study on the identification of QTL involved in the control of heterosis in maize, Frascaroli et al. (2007) observed that QTL for traits with low heterosis were prevailingly in the additive to a dominance range, while QTL for highly heterotic traits were characterized by effects in the dominance to overdominance range. However, Shang et al. (2016), while studying the genetic basis of heterosis in Upland cotton using two connected RIL populations and two corresponding BCF_1 populations, observed that partial dominance and epistasis played a relatively more important role than other genetic effects in heterosis in Upland cotton. Studies on maize (Stuber et al., 1992) and (Frascaroli et al., 2007) have also highlighted the role of overdominance in heterosis. Overdominance has also been identified as the primary genetic basis of heterosis in rice (Li et al., 2001; Luo et al., 2001) and tomato (Krieger et al., 2010). The results in this study show

that, although all levels of dominance effects are responsible for heterosis, overdominance is the major contributing factor in the heterosis in *B. juncea* (**Table 3**). Similar results were reported by Radoev et al. (2008) in *B. napus*, in which the phenotypic variances explained by 14 QTL, showing overdominance, were much larger than that shown by 13 QTL, exhibiting partial- to full-dominance loci for the nine heterotic traits.

Previous studies on *B. napus* (Quijada et al., 2006; Udall et al., 2006; Chen et al., 2007; Basunanda et al., 2010) have also identified heterosis-related QTL clusters, influencing yield-related traits in different DH and/or testcross populations. The active region mapped in the marker interval Bn-A07-p21095697-Bn-A07-p18983802 in linkage group A07 was consistently detected as major QTL in individual trials as well as using the BLUPs, indicating the authentic nature of this QTL and, therefore, should be a focus for further in-depth studies on yield-related traits in *B. juncea* and related species.

Whole-Genome Heterozygosity and Trait Performance

It was observed in the present study that not all the traits showed higher phenotypes in the backcross lines (BC-V and BC-E heterozygotes) than in their respective DH (VEH homozygotes). QTL detected in both the MPH data sets were a mixture of overdominant and underdominant loci, connoting heterozygosity was not always necessarily advantageous for the expression of the trait. Although significant correlations of whole-genome heterozygosity were observed with some of the traits in the backcross populations and with their respective MPH data sets, the magnitudes of correlations were low (**Table 2**). It, therefore, appears that heterosis in *B. juncea* arises mainly from heterozygosity at specific loci rather than whole-genome heterozygosity. Similar results were found previously in rice (Yu et al., 1997; Hua et al., 2002; Mei et al., 2005; Luo et al., 2009), upland cotton (Shang et al., 2016), and rapeseed (Radoev et al., 2008). Evidently, the low correlations could also exist due to sparsity of markers or a 50% reduced representation of heterozygous loci in backcross progenies, but, with a higher density of markers and increased genome coverage in the VEH map, the results in this study could be more accurate than other studies, which lacked these features of genetic maps. Large-scale genome-wide association study (GWAS) by Huang et al. (2015) investigated population-scale genomic landscapes from a representative number of hybrid varieties of rice (a predominantly self-pollinated crop) along with parental lines and showed that overall heterozygosity made only small contribution to heterosis. It was concluded that the accumulation of several rare superior alleles with positive dominance contributes to the expression of heterosis in the rice hybrids.

Mapping QTL With Epistatic Effects

A large number of epistatic interactions were detected in the present study, indicating that epistasis plays an important role in the expression of heterosis in *B. juncea*. The results follow the findings in previous studies on heterosis in rapeseed (Radoev et al., 2008) and Upland cotton (Shang et al., 2016) in which a large number of digenic epistatic interactions were also identified. In studies conducted on maize by Stuber et al. (1992), Lu et al. (2003), and Mihaljevic et al. (2005), and on rice by Xiao et al. (1995), no significant epistasis was detected. In Arabidopsis, Kusterer et al. (2007) found a significant role of epistasis in the expression of heterosis for five biomass-related traits in a triple-testcross design with a recombinant inbred line population. Melchinger et al. (2007) also confirmed the role of epistasis in the expression of heterosis in Arabidopsis, using NIL libraries.

In this study, out of 637 epistatic interactions detected, the vast majority of 532 (83.5%) interactions belonged to the type III class, while 102 (16%) interactions were identified as type II. Only three (0.5%) type I interactions were detected collectively for all the 14 traits using the five data sets (**Table 4**). In several studies on understanding the genetic basis of heterosis in autogamous species like Arabidopsis (Kusterer et al., 2007; Melchinger et al., 2007), rice (Yu et al., 1997; Li et al., 2001, 2008; Luo et al., 2001; Mei et al., 2005) and barley (Xu and Jia, 2007), epistasis between background loci has been shown to play a major role in comparison to the contribution of loci with main effects, whereas loci with main effects have been ascribed a principal role in determining heterosis in autogamous species such as maize (Stuber et al., 1992; Lu et al., 2003; Frascaroli et al., 2007; Tang et al., 2010; Yi et al., 2019; Xiao et al., 2021). The difference in the genetic basis of heterosis, therefore, appears to be associated with the mode of pollination – open or self-pollination. Radoev et al. (2008), Basunanda et al. (2010), Shi et al. (2011) and Li et al. (2012) have also reported the predominance of epistatic interactions in influencing heterosis in *B. napus*. It has been hypothesized that, in autogamous species, positive epistatic effects due to coadapted gene complexes can be perpetuated easily, compared with allogamous species (Allard, 1988; Garcia et al., 2008; Shi et al., 2011). As both *B. juncea* (5–20%) and *B. napus* (10–30%) are partially allogamous crops (Asthana and Singh, 1973; Rakow and Woods, 1987), the important role played by epistasis in regulating the genetic mechanisms underlying heterosis appears to be similar (Radoev et al., 2008; Basunanda et al., 2010; Shi et al., 2011; Li et al., 2012). The fact that, in this study, most of the epistasis occurred between complementary loci without detectable main effects is found to be in sharp agreement with the results reported by Radoev et al. (2008), Basunanda et al. (2010), and Shi et al. (2011).

A number of loci involved in epistasis were found to be interacting with more than one locus. For example, in BC-V population, the locus on A7 involved in plant height interacted with loci on B1 and B6. Such interactions were also observed where one locus interacted with two different loci, which were detected in different data sets. For example, a locus detected on linkage group B1 (39.48 cM to 40.7 cM) interacted with a locus on linkage group B6 (59.85 cM to 60.08 cM) detected in BC-E and also with a locus on linkage group B2 (167.87 cM to 170.43 cM) detected in MPH-V. This participation of loci in multi-locus interactions has been hypothesized to constitute higher order epistatic interactions (Zhao et al., 2005; Radoev et al., 2008). Similar results were observed by Shang et al. (2016) in upland cotton, where they found E-QTL, showing pleiotropic effects.

This research revealed that the genetic basis of heterosis of yield-influencing traits in *B. juncea* is complicated but definitely

involves a large number of loci-exhibiting cumulative effects of dominance, overdominance, and a large number of epistatic interactions. The QTL detected in multiple environments can be considered as candidate loci for further studies. This study will provide the essential stimulus for further research on heterosis studies on *B. juncea*, which can be performed by evaluating testcrosses with related or unrelated testers to uncover stable QTL in inbred lines and across testers. The selection of these consistent QTL across testers can lead to better yield performance and in the development of hybrid combinations. The common QTLs validated across multiple trials and multiple populations will provide a valuable resource for MAS and the further research.

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/s.

AUTHOR CONTRIBUTIONS

Aakanksha, and SY carried out the phenotyping and mapping work. BY helped in field experiments and candidate gene analysis. VG helped with genotyping. AM developed microspore-derived F_1DH mapping population. DP and AP conceived and supervised the overall study and along with Aakanksha and SY wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

This study was funded by the Department of Biotechnology (DBT), Government of India through the award of two projects—(1) Centre of Excellence on Genome Mapping and Molecular Breeding of Brassicas (BT/01/COE/08/06/-II) and (2) DBTUDSC Partnership Centre on Genetic Manipulation of Brassicas (BT/01/NDDB/UDSC/2016).

ACKNOWLEDGMENTS

Support of the National Dairy Development Board (NDDB) for the research till September, 2016 is acknowledged. Aakanksha acknowledges the receipt of research fellowship from the Council of Scientific and Industrial Research (CSIR). DP was supported with a JC Bose fellowship from the Department of Science & Technology (DST) and CSIR as a distinguished scientist.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.721631/full#supplementary-material

REFERENCES

Allard, R. W. (1988). Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. *J. Hered.* 79, 225–238. doi: 10.1093/oxfordjournals.jhered.a110503

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant. Mol. Biol. Rep.* 9, 208–218. doi: 10.1007/BF02672069

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Asthana, A. N., and Singh, C. B. (1973). Hybrid vigour in rai. *Indian. J. Genet. Plant Breed.* 33, 57–63.

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

Cai, G., Yang, Q., Chen, H., Yang, Q., Zhang, C., Fan, C., et al. (2016). Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Rep.* 6:21625. doi: 10.1038/srep21625

Chen, W., Zhang, Y., Liu, X., Chen, B., Tu, J., and Tingdong, F. (2007). Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F_2 populations. *Theor. Appl. Genet.* 115, 849–858. doi: 10.1007/s00122-007-0613-2

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.

Clarke, W. E., Higgins, E. E., Plieske, J. Wieseke, R., Sidebottom, C., Khedikar, Y., Batley, J., et al. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899. doi: 10.1007/s00122-016-2746-7

Cockerham, C. C., and Zeng, Z. B. (1996). Design III with marker loci. *Genetics* 143, 1437–1456.

Crow, J. F. (1999). "Dominance and overdominance," in *Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Madison, WI: American Society of Agronomy).

Desveaux, D., Marechal, A., and Brisson, N. (2005). Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci.* 10, 95–102. doi: 10.1016/j.tplants.2004.12.008

Dhaka, N., Mukhopadhyay, A., Paritosh, K., Gupta, V., Pental, D., and Pradhan, A. K. (2017). Identification of genic SSRs and construction of a SSR-based linkage map in *Brassica juncea*. *Euphytica* 213:15. doi: 10.1007/s10681-016-1814-z

East, E. M. (1908). Inbreeding in corn. *Rep. Conn. Agric. Exp. Stn.* 1907–1908, 419–428.

Frascaroli, E., Canè, M. A., Landi, P., Pea, G., Gianfranceschi, L., Villa, M., et al. (2007). Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* 176, 625–644. doi: 10.1534/genetics.106.064493

Frascaroli, E., Canè, M. A., Pè, M. E., Pea, G., Morgante, M., and Landi, P. (2009). QTL detection in maize testcross progenies as affected by related and unrelated testers. *Theor. Appl. Genet.* 118, 993–1004. doi: 10.1007/s00122-008-0956-3

Fu, J., Keurentjes, J., Bouwmeester, H., America, T., Verstappen, F.W., Ward, J.L., et al. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat. Genet.* 41, 166–167. doi: 10.1038/ng.308

Garcia, A. A. F., Wang, S., Melchinger, A. E., and Zeng, Z. B. (2008). Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics* 180: 1707–1724. doi: 10.1534/genetics.107.082867

Goodnight, C. J. (1999). "Epistasis and heterosis," in *Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Madison, WI: American Society of Agronomy). doi: 10.2134/1999.geneticsandexploitation.c6

Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.* 57, 963–972. doi: 10.3168/jds.S0022-0302(74)84993-3

Hua, J., Xing, Y., Wu, W., Xu, C., Sun, X., Yu, S., et al. (2003). Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2574–2579. doi: 10.1073/pnas.0437907100

Hua, J., Xing, Y., Xu, C., Sun, X., Yu, S., and Zhang, Q. (2002). Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics* 162, 1885–1895.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6, 1–9. doi: 10.1038/ncomms7258

Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene single flower truss drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463. doi: 10.1038/ng.550

Kusterer, B., Muminovic, J., Utz, H. F., Piepho, H. P., Barth, S., Heckenberger, M., et al. (2007). Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in *Arabidopsis*. *Genetics* 175, 2009–2017. doi: 10.1534/genetics.106.069005

Larièpe, A., Mangin, B., Jasson, S., Combes, V., Dumas, F., Jamin, P., et al. (2012). The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). *Genetics* 190, 795–811. doi: 10.1534/genetics.111.133447

Li, C., Zhao, T., Yu, H., Li, C., Deng, X., Dong, Y., et al. (2018). Genetic basis of heterosis for yield and yield components explored by QTL mapping across four genetic populations in upland cotton. *BMC Genomics* 19:910. doi: 10.1186/s12864-018-5289-2

Li, L., Lu, K., Chen, Z., Mu, T., Hu, Z., and Li, X. (2008). Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics* 180, 1725–1742. doi: 10.1534/genetics.108.091942

Li, Y., Zhang, X., Ma C., Shen, J., Chen, Q., Wang, T., et al. (2012). QTL and epistatic analyses of heterosis for seed yield and three yield component traits using molecular markers in rapeseed (*Brassica napus L.*). *Genetika* 48, 1171–1178.

Li, Z. K., Luo, L. J., Mei, H. W., Wang, D. L., Shu, Q. Y., Tabien, R., et al. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* 158, 1737–1753. doi: 10.3410/f.1002133.8203

Liang, Q., Shang, L., Wang, Y., and Hua, J. (2015). Partial dominance, overdominance and epistasis as the genetic basis of heterosis in upland cotton (*Gossypium hirsutum* L.). *PLoS ONE* 10:e0143548. doi: 10.1371/journal.pone.0143548

Lippman, Z. B., and Zamir, D. (2007). Heterosis: revisiting the magic. *Trends Genet.* 23, 60–66. doi: 10.1016/j.tig.2006.12.006

Liu, H., Wang, Q., Chen, M., Ding, Y., Yang, X., Liu, J., et al. (2020). Genome-wide identification and analysis of heterotic loci in three maize hybrids. *Plant Biotechnol. J.* 18,185–194. doi: 10.1111/pbi.13186

Lu, H., Romero-Severson, J., and Bernardo, R. (2003). Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. *Theor. Appl. Genet.* 107, 494–502. doi: 10.1007/s00122-003-1271-7

Luo, L.J., Li, Z.K., Mei, H.W., Shu, Q.Y., Tabien, R., Zhong, D.B., et al. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* 158, 1755–1771.

Luo, X., Fu, Y., Zhang, P., Wu, S., Tian, F., Liu, J., et al. (2009). Additive and over-dominant effects resulting from epistatic loci are the primary genetic basis of heterosis in rice. *J. Integr. Plant. Biol.* 51, 393–408. doi: 10.1111/j.1744-7909.2008.00807.x

Ma, L., Wang, Y., Ijaz, B., and Hua, J. (2019). Cumulative and different genetic effects contributed to yield heterosis using maternal and paternal backcross populations in Upland cotton. *Sci. Rep.* 9:3984. doi: 10.1038/s41598-019-40611-9

Mason, A. S., Higgins, E. E., Snowdon, R. J., Batley, J., Stein, A., Werner, C., et al. (2017). A user guide to the Brassica 60K Illumina Infinium^{TM} SNP genotyping array. *Theor. Appl. Genet.* 130, 621–633. doi: 10.1007/s00122-016-2849-1

Mei, H. W., Li, Z. K., Shu, Q. Y., Guo, L. B., Wang, Y. P., Yu, X. Q., et al. (2005). Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. *Theor. Appl. Genet.* 110, 649–659. doi: 10.1007/s00122-004-1890-7

Mei, H. W., Luo, L. J., Ying, C. S., Wang, Y. P., Yu, X. Q., Guo, L. B., et al. (2003). Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. *Theor. Appl. Genet.* 107, 89–101. doi: 10.1007/s00122-003-1192-5

Melchinger, A. E., Piepho, H. P., Utz, H. F., Muminovic, J., Wegenast, T., Torjek, O., et al. (2007). Genetic basis of heterosis for growth-related traits in *Arabidopsis* investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics* 177, 1827–1837. doi: 10.1534/genetics.107.080564

Melchinger, A. E., Utz, H. F., and Schön, C. C. (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149, 383–403. doi: 10.1093/genetics/149.1.383

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., and Thomas, P. D. (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* 38, D204–D210. doi: 10.1093/nar/gkp1019

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., et al. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49, D394–D403. doi: 10.1093/nar/gkaa1106

Mihaljevic, R., Utz, H. F., and Melchinger, A. E. (2005). No evidence for epistasis in hybrid and per se performance of elite European flint maize inbreds from generation means and QTL analyses. *Crop Sci.* 45, 2605–2613. doi: 10.2135/cropsci2004.0760

Mukhopadhyay, A., Arumugam, N., Sodhi, Y., Gupta, V., Pradhan, A., and Pental, D. (2007). "High frequency production of microspore derived doubled haploid (DH) and its application for developing low glucosinolate lines in Indian *Brassica juncea*," in *Proceedings of the 12th International Rapeseed Congress*, 333–335.

Olivoto, T., and Lúcio, A. D. C. (2020). Metan: An R package for multi-environment trial analysis. *Methods Ecol. Evol.* 11, 783–789. doi: 10.1111/2041-210X.13384

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

Paritosh, K., Gupta, V., Yadava, S. K., Singh, P., Pradhan, A. K., and Pental, D. (2014). RNA-seq based SNPs for mapping in *Brassica juncea* (AABB): synteny analysis between the two constituent genomes A (from *B. rapa*) and B (from *B. nigra*) shows highly divergent gene block arrangement and unique block fragmentation patterns. *BMC Genomics* 15:396. doi: 10.1186/1471-2164-15-396

Paritosh, K., Yadava, S. K., Singh, P., Bhayana, L., Mukhopadhyay, A., Gupta, V., et al. (2021). A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *Plant Biotechnol. J.* 19:602. doi: 10.1111/pbi.13492

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726. doi: 10.1038/335721a0

Peng, J., Ronin, Y., Fahima, T., Röder, M. S., Li, Y., Nevo, E., et al. (2003). Domestication quantitative trait loci in *Triticum dicoccoides,* the progenitor of wheat. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2489–2494. doi: 10.1073/pnas.252763199

Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A., et al. (1999). Mining SNPs from EST databases. *Genome Res.* 9, 167–174. doi: 10.1101/gr.9.2.167

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing, *Euphytica* 161, 209–228. doi: 10.1007/s10681-007-9449-8

Pradhan, A. K., and Pental, D. (2011). "Genetics of *Brassica juncea* L," in *Genetics and Genomics of the Brassicaceae*, eds R. Schmidt and I. Bancroft (New York, NY: Springer), 323–346. doi: 10.1007/978-1-4419-7118-0_11

Pradhan, A. K., Sodhi, Y. S., Mukhopadhyay, A., and Pental, D. (1993). Heterosis breeding in Indian mustard (*Brassica juncea* L. Czern and Coss): analysis of component characters contributing to heterosis for yield. *Euphytica* 69, 219–229. doi: 10.1007/BF00022368

Quijada, P. A., Udall, J. A., Lambert, B., and Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theor. Appl. Genet.* 113, 549–561. doi: 10.1007/s00122-006-0323-1

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radoev, M., Becker, H. C., and Ecke, W. (2008). Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics* 179, 1547–1558. doi: 10.1534/genetics.108.089680

Rakow, G., and Woods, D. L. (1987). Outcrossing in rape and mustard under Saskatchewan prairie conditions. *Can J. Plant Sci.* 67, 141–151. doi: 10.4141/cjps87-017

Ramchiary, N., Padmaja, K. L., Sharma, S., Gupta, V., Sodhi, Y. S., Mukhopadhyay, A., et al. (2007). Mapping of yield influencing QTL in *Brassica juncea*: implications for breeding of a major oilseed crop of dryland areas. *Theor. Appl. Genet.* 115, 807–817. doi: 10.1007/s00122-007-0610-5

Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratgé-Faillie, C., Novák, O., et al. (2013). Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homoeostasis. *Nat. Commun.* 4:2625. doi: 10.1038/ncomms3625

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., et al. (2000). *Arabidopsis* transcription factors: genome wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105

Rogers, S. O., and Bendich, A. J. (1994). "Extraction of total cellular DNA from plants, algae and fungi," in *Plant Molecular Biology Manual*, eds S. B. Gelvin and R. A. Schilperoort (Dordrecht: Springer), 1–8. doi: 10.1007/978-94-011-0511-8_12

Rout, K., Yadav, B. G., Yadava, S. K., Mukhopadhyay, A., Gupta, V., Pental, D., and Pradhan, A. K. (2018). QTL landscape for oil content in *Brassica juncea*: Analysis in multiple bi-parental populations in high and "0" erucic background. *Front. Plant Sci.* 9:1448. doi: 10.3389/fpls.2018.01448

Schnable, P. S., and Springer, N. M. (2013). Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64, 71–88. doi: 10.1146/annurev-arplant-042110-103827

Semel, Y., Nissenbaum, J., Menda, N., Zinder, M., Krieger, U., Issman, N., et al. (2006). Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12981–12986. doi: 10.1073/pnas.0604635103

Shang, L., Wang, Y., Cai, S., Wang, X., Li, Y., Abduweli, A., et al. (2016). Partial dominance, overdominance, epistasis and QTL by environment interactions contribute to heterosis in two upland cotton hybrids. *Genes Genomes Genetics* 6, 499–507. doi: 10.1534/g3.115.025809

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

Shull, G. H. (1908). The composition of a field of maize. *J. Hered.* 1, 296–301.

Sodhi, Y.S., Chandra, A., Verma, J. K., Arumugam, N., Mukhopadhyay, A., Gupta, V., et al. (2006). A new cytoplasmic male sterility system for hybrid seed production in Indian oilseed mustard *Brassica juncea*. *Theor. Appl. Genet*. 114, 93–99. doi: 10.1007/s00122-006-0413-0

Song, J., Li, B., Cui, Y., Zhuo, C., Gu, Y., Hu, K., et al. (2021). QTL mapping and diurnal transcriptome analysis identify candidate genes regulating *Brassica napus* flowering time. *Int. J. Mol. Sci.* 22:7559. doi: 10.3390/ijms22147559

Srivastava, A., Gupta, V., Pental, D., and Pradhan, A. K. (2001). AFLP-based genetic diversity assessment amongst agronomically important natural and some newly synthesized lines of *Brassica juncea*. *Theor. Appl. Genet.* 102, 193–199. doi: 10.1007/s001220051635

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi: 10.1016/S1369-5266(00)00199-0

Stuber, C. W. (1992). Biochemical and molecular markers in plant breeding. *Plant Breed. Rev.* 9, 37–61.

Stuber, C. W., Edwards, M. D. A., and Wendel, J. F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Sci.* 27, 639–648. doi: 10.2135/cropsci1987.0011183X002700040006x

Stuber, C. W., Lincoln, S. E., Wolff, D., Helentjaris, T., and Lander, E. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132, 823–839. doi: 10.1093/genetics/132.3.823

Swanson-Wagner, R. A., DeCook, R., Jia, Y., Bancroft, T., Ji, T., Zhao, X., et al. (2009). Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. *Science* 326, 1118–1120. doi: 10.1126/science.1178294

Tang, J., Yan, J., Ma, X., Teng, W., Wu, W., Dai, J., et al. (2010). Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. *Theor. Appl. Genet.* 120, 333–340. doi: 10.1007/s00122-009-1213-0

Taylor, J., and Butler, D. (2017). R Package ASMap: efficient genetic linkage map construction and diagnosis. *J. Stat. Softw.* 79, 1–29. doi: 10.18637/jss.v079.i06

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13: 2129–2141. doi: 10.1101/gr.772403

Udall, J. A., Quijada, P. A., Lambert, B., and Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 2. Identification of alleles from unadapted germplasm. *Theor. Appl. Genet.* 113, 597–609. doi: 10.1007/s00122-006-0324-0

Utz, H. F. (2004). *PLABSTAT. A Computer Program for Statistical Analysis of Plant Breeding Experiments*. Inst. of Plant Breeding, Seed Sci., and Population Genetics, Univ. of Hohenheim, Stuttgart, Germany.

Voorrips, R. E. (2002). MapChart. Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Vroemen, C. W., Mordhorst, A. P., Albrecht, C., Kwaaitaal, M. A., and de Vries, S. C. (2003). The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in *Arabidopsis*. *Plant Cell* 15, 1563–1577. doi: 10.1105/tpc.012203

Wang, C. L., Zhang, Q., Jiang, L., Qian, R., Ding, X., and Zhao, Y. (2016). Comparative study of estimation methods for genomic breeding values. *Sci. Bull.* 61, 353–356. doi: 10.1007/s11434-016-1014-1

Wang, H., Zaman, Q. U., Huang, W., Mei, D., Liu, J., Wang, W., et al. (2019). QTL and candidate gene identification for silique length based on high-dense genetic map in *Brassica napus* L. *Front. Plant Sci.* 10:1579. doi: 10.3389/fpls.2019.01579

Wang, S., Basten, C., and Zeng, Z. (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Wu, J., Chen, P., Zhao, Q., Cai, G., Hu, Y., Xiang, Y., et al. (2018). Co-location of QTL for *Sclerotinia* stem rot resistance and flowering time in *Brassica napus*. *Crop J.* 7, 227–237. doi: 10.1016/j.cj.2018.12.007

Xiao, J. H., Li, J. M., Yuan, L. P., and Tanksley, S. D. (1995). Dominance is the major genetic-basis of heterosis in rice as revealed by Qtl analysis using molecular markers. *Genetics* 140, 745–754. doi: 10.1093/genetics/140.2.745

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* 22:148. doi: 10.1186/s13059-021-02370-7

Xu, S., and Jia, Z. (2007). Genomewide analysis of epistatic effects for quantitative traits in barley. *Genetics* 175, 1955–1963. doi: 10.1534/genetics.106.066571

Yadava, S. K., Arumugam, N., Mukhopadhyay, A., Sodhi, Y. S., Gupta, V., Pental, D., et al. (2012). QTL mapping of yield-associated traits in *Brassica juncea*: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. *Theor. Appl. Genet.* 125, 1553–1564. doi: 10.1007/s00122-012-1934-3

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* 48, 1225–1232. doi: 10.1038/ng.3657

Yang, Y., Shen, Y., Li, S., Ge, X., and Li, Z. (2017). High Density Linkage Map Construction and QTL detection for three silique-related traits in *Orychophragmus violaceus* derived *Brassica napus* population. *Front. Plant Sci.* 8:1512. doi: 10.3389/fpls.2017.01512

Yi, Q., Liu, Y., Hou, X., Zhang, X., Li, H., Zhang, J., et al. (2019). Genetic dissection of yield-related traits and mid-parent heterosis for those traits in maize (Zea mays L.). *BMC Plant Biol.* 19:392. doi: 10.1186/s12870-019-2009-2

You, A., Lu, X., Jin, H., Ren, X., Liu, K., Yang, G., et al. (2006). Identification of quantitative trait loci across recombinant inbred lines and testcross populations for traits of agronomic importance in rice. *Genetics* 172, 1287–1300. doi: 10.1534/genetics.105.047209

Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., et al. (1997). Importance of epistasis as the genetic basis of hete rosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9226–9231. doi: 10.1073/pnas.94.17.9226

Zhao, H., Yan, W., Yu, K., Wang, T., Khattak, A. N., and Tian, E. (2021). QTL identification for nine seed-related traits in *Brassica juncea* using a multiparent advanced generation intercross (MAGIC) population. *Czech J. Genet. Plant Breed.* 57, 9–18. doi: 10.17221/73/2020-CJGPB

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

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

Zhihong, Z., Yousaf, H., Jian, Y., Liyong, C., Xiangyang, L., and Haiming, X. (2012). Statistical method for mapping QTLs for complex traits based on two backcross populations. *Chin. Sci. Bull.* 57, 2645–2654. doi: 10.1007/s11434-012-5279-8

Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., et al. (2012). Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15847–15852. doi: 10.1073/pnas.1214141109

Zhou, J. J., and Luo, J. (2018). The PIN-FORMED auxin efflux carriers in plants. *Int. J. Mol. Sci.* 19:2759. doi: 10.3390/ijms19092759

Zhu, Y. J., Huang, D. R., Fan, Y. Y., Zhang, Z. H., Ying, J. Z., and Zhuang, J. Y. (2016). Detection of QTLs for yield heterosis in rice using a RIL population and its testcross population. *Int J. Genomics* 2016:2587823. doi: 10.1155/2016/2587823

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 in Plant Science

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fpls.2021.726084

Genome-Wide Analysis of Simple Sequence Repeats in Cabbage (*Brassica oleracea* L.)

Yuanyuan Xu, Miaomiao Xing, Lixiao Song, Jiyong Yan, Wenjiang Lu and Aisong Zeng**

Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Cabbage (*Brassica oleracea* L. var. *capitata*) accounts for a critical vegetable crop belonging to Brassicaceae family, and it has been extensively planted worldwide. Simple sequence repeats (SSRs), the markers with high polymorphism and co-dominance degrees, offer a crucial genetic research resource. The current work identified totally 64,546 perfect and 93,724 imperfect SSR motifs in the genome of the cabbage 'TO1000.' Then, we divided SSRs based on the respective overall length and repeat number into different linkage groups. Later, we characterized cabbage genomes from the perspectives of motif length, motif-type classified and SSR level, and compared them across cruciferous genomes. Furthermore, a large set of 64,546 primer pairs were successfully identified, which generated altogether 1,113 SSR primers, including 916 (82.3%) exhibiting repeated and stable amplification. In addition, there were 32 informative SSR markers screened, which might decide 32 cabbage genotypes for their genetic diversity, with level of polymorphism information of 0.14–0.88. Cultivars were efficiently identified by the new strategy designating manual diagram for identifying cultivars. Lastly, 32 cabbage accessions were clearly separately by five Bol-SSR markers. Besides, we verified whether such SSRs were available and transferable in 10 Brassicaceae relatives. Based on the above findings, those genomic SSR markers identified in the present work may facilitate cabbage research, which lay a certain foundation for further gene tagging and genetic linkage analyses, like marker-assisted selection, genetic mapping, as well as comparative genomic analysis.

Keywords: cabbage, genome, SSR, molecular makers, genetic diversity, manual cultivar identification diagram

OPEN ACCESS

Edited by:
Ryo Fujimoto,
Kobe University, Japan

Reviewed by:
Angelica Giancaspro,
University of Bari Aldo Moro, Italy
Zhou Xirong,
Shanghai Academy of Agricultural Sciences, China

****Correspondence:***
Jiyong Yan
yanjy0266@126.com
Aisong Zeng
zengaisong1960@163.com

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

Received: *16 June 2021*
Accepted: *15 November 2021*
Published: *09 December 2021*

Citation:
Xu Y, Xing M, Song L, Yan J, Lu W and Zeng A (2021) Genome-Wide Analysis of Simple Sequence Repeats in Cabbage (Brassica oleracea L.). Front. Plant Sci. 12:726084. doi: 10.3389/fpls.2021.726084

INTRODUCTION

Molecular markers used for research change from enzyme-based to DNA-based ones at present; moreover, numerous DNA markers systems are constructed. (Zhang et al., 2016; Ikten et al., 2019). Simple sequence repeats (SSRs), which are also referred to as microsatellites, represent the tandem repeat sequences containing 2–6 nucleotides (nt) short units and are usually seen in eukaryotic and prokaryotic genomes. SSRs are identified to be the preferred option for different research (Wang et al., 2021; Zhang et al., 2021). SSR markers are highly variable, abundant, reproducible, and transferrable, with co-dominant and multi-allelic inheritance; as a result, they are the precious and creditable approaches to carry out gene tapping, genetic mapping, comparative mapping, and genetic diversity analyses on plant species (Li et al., 2012;

Silva et al., 2013). Generally, SSR markers are developed dependent on SSR motifs as well as the corresponding flanking sequences, and they may be separated from non-coding nt sequences or conserved coding regions in each higher organism (Sraphet et al., 2011; Xu et al., 2019). Cross-species amplification has been conducted to discover SSR markers that can be used in plant research, and it is related to the selection of genomic libraries or SSR-abundant cDNA and the search of open databases (Yang et al., 2020). At present, the whole-genome sequence for one specific species can be available, which allows to identify and develop SSR markers at genome-wide level (Karci et al., 2020). Notably, the emergence of next-generation sequencing (NGS) technology has reduced the time and cost necessary to carry out whole-genome sequencing (WGS) for plant species (Portis et al., 2018). Genomes make it possible for the development of numerous SSR markers for assessing genetic variations in germplasms and cultivars, identifying quantitative trait loci and genes for the control of traits with economic importance, develop molecular genetics and physical maps and assist in breeding for crop improvement (Cui et al., 2017; Xue et al., 2018). SSR markers have been identified in the whole-genome of diverse living bodies, such as human beings, insects, marine animals, plants with economic value and medicinal fungi (Gil et al., 2017; Liu et al., 2018, 2019). Nevertheless, with the exception of Chinese cabbage, eggplant and cucumber, research on key vegetable species like cabbage in this field is lacking.

Cabbage (*Brassica oleracea* L., $2n = 18$) is one of the most critical cruciferous vegetables that is widely cultivated all over the world. As a kind of vegetable, cabbage has been extensively consumed throughout the world because it contains favorable components for human health (Lv et al., 2014; Cai et al., 2020). Some molecular markers are used within cabbage, like sequence-related amplified polymorphism (SRAP), universal random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism and amplified fragment length polymorphism (AFLP) markers (Ikten et al., 2019). Currently, informative molecular markers have emerged, which offer precious knowledge for genomic and genetic research on cabbage (Gadaleta et al., 2012; Zhong et al., 2017). However, regardless of the development of SSR markers, those developed from cabbage reference genome are lacking compared with those in other crops. There are inadequate SSR markers to construct linkage mapping or association research on cabbage (Lee et al., 2015; Lv et al., 2017). Additionally, prior works that develop SSR markers for cabbage mostly focus on screening SSR markers from public databases or SSR-abundant libraries (Taheri et al., 2018). The plenty of RNA-Seq data have contributed to *in silico* generation, but there is no available comprehensive analysis on SSRs within cabbage genome, even though related information becomes accessible recently. Moreover, SSR loci have been increasingly utilized for the development of molecular markers that can be applied in genetic analysis, like genome assembly, positional cloning, diversity assessment, as well as breeding activities like marker-assisted selection (MAS), but they are not used or detected in further research.

The release of reference genome of cabbage has produced numerous sequences, which are adopted to develop and identify cabbage SSR markers (Cai et al., 2020; Lv et al., 2020). Thus, the present work focused on identifying SSRs at genome-wide level in the cabbage line TO1000, a homozygous doubled haploid, and evaluate them for marker development.

A large set of 64,546 primer pairs was successfully identified in the cabbage genome. Among these, there were altogether 1,113 SSR primers prepared, and a subset of 916 (82.3%) pairs of primers could be stably and repeatably amplified. Moreover, according to the genetic diversity analyses on 32 cabbage genotypes, we identified a set of 32 SSR markers. By adopting the manual cultivar identification diagram (MCID), it was possible to rapidly distinguish cabbage genotypes by combining 5 SSR primers, the novel strategy enabling the practical and referable application of molecular markers and morphological descriptors. Additionally, SSRs detected within the conserved coding regions were highly available and transferable among 10 relevant species of cruciferous crops; as a result, they were conductive to comparative analysis on relatives belonging to Brassicaceae family. Data on SSR markers contribute to the rapid enrichment of functional molecular markers closely associated with expressed regions within cabbage genes, which show high value in the comparative genomic analysis and genetic mapping of cabbage.

MATERIALS AND METHODS

Plant Materials

A total of 32 cabbage genotypes (**Supplementary Table S1**) that had diverse morphologies and origins were chosen to analyze the genetic diversity and identify the cultivars, besides, 10 relatives were also chosen from Brassicaceae to study the transferability.

SSR Content of the Cabbage Genome

This study obtained the high-quality cabbage genome within the homozygous doubled haploid "TO1000," in the format of FASTA (freely accessible at www.ncbi.nlm.nih.gov/genome/10901). Thereafter, we cut 9 pseudomolecules that stood for part of chromosomal sequences in every species, together with those unmapped scaffolds, to small pieces by adopting SciRoKo tool.[1] Later, SciRoKo SSR-search module was adopted for the *in silico* identification of imperfect, perfect, and compound SSRs. Search queries were specified as at least 4 repetitions and at least 15 nt in length. Perfect SSR was defined as a sequence in which one motif was repeated for 4 times (4–6 nt motif), 5 times (3 nt), 8 times (2 nt), and 15 times (1 nt), with just one mismatch. As for compound repeats, we set the maximal length of default interruption (spacer) as 100 bp. Thereafter, Bedtools was adopted to match those coordinates (start/end positions) for every SSR with gene space, for the sake of intersecting with default parameters under the left outer join option. A repeat was called an SSR when there

[1]http://kofler.or.at/bioinformatics/SciRoKo

was one or more than 1 nt in the overlap. GO analysis was conducted to define possible gene function carrying one or more SSRs. The enriched GO terms were examined through analyzing SSRs set in genome-wide GO annotation dataset by R ClusterProfiler (v. 3.6.0) package, collected upon the thresholds of false discovery rate < 0.01 and values of $p < e^{-5}$, and visualized using Cytoscape v. 3.7.1.

Collection of Genomic Sequences From Different Cruciferous Crops

To compare, we obtained genome sequences for homozygous doubled haploid cabbage "TO1000," and full-genome sequences for 11 additional plant cultivars in Brassicaceae family, including *Brassica napus*, *Brassica rapa* subsp. *pekinensis*, *Arabidopsis thaliana*, *Brassica cretica, Raphanus sativus*, *Brassica nigra, Brassica juncea*, *Camelina sativa*, *Capsella bursa-pastoris*, *Eutrema yunnanense*, *Barbarea vulgaris*, and *B. oleracea*, based on open database. Then, we performed the above-mentioned process to scan whether perfect SSRs existed. **Supplementary Table S2** displays the sources of all the full-genome sequences.

SSR Identification and Primer Design

SSRs were identified in whole-genome data of cabbage using the MISA package. The SSR motif length was restrained to 1–6 bp, which was in consistence with mononucleotides (Mono-), dinucleotides (Di-), trinucleotides (Tri-), tetranucleotides (Tetra-), pentanucleotides (Penta-), and hexanucleotides (Hexa-), separately. Search standards were the same as those in previous study (Cheng et al., 2016). In addition, primers were designed by adopting Perl scripts p3_in.pl./p3_out.pl.[2] and Primer3 primer modeling software,[3] and SSR search findings were used to be the input.

Conditions to select primers were shown below, primer size, 18–27 bp (optimal, 20 bp); melting temperature of primer (*T*m), 57.0–63.0°C (optimal, 60°C), primer GC level, 40–60% (best, 50%) and product size, 100–500 bp (optimal, 250 bp). Each of the primer pairs designed was later aligned against the 'TO1000' cabbage reference genome. We defined unique primer pairs as those whose reverse and forward primers showed unique alignment to reference genome with a 100% match rate.

DNA Extraction, PCR Amplification and Detection

We chose altogether 48 primer pairs at random for better validating amplification of the particular SSR primer set identified in the present work (**Supplementary Table S3**). Genomic DNA (gNDA) fragments of 'QBYS' and 'QBJF' cabbage lines were amplified by using every pair of primers. The CTAB protocol after modification (Liu et al., 2003) was also employed to extract gDNA from the young leaf samples in 32 cabbage accessions of diverse origins and in 10 relevant *Brassica* species.

The 20 μl volume was prepared for every PCR procedure, including template DNA (10 ng), $MgCl_2$ (2.0 mm), dNTPs (0.2 mm), respective primers (0.1 μm) and Taq DNA polymerase (0.5 U, TaKaRa Bio Inc., Dalian, China). The reaction procedure was as follows, 3 min of initial denaturation under 94°C; 50 s under 94°C, 50 s under 56°C, as well as 1 min under 72°C for 35 cycles; final 10 min of extension under 72°C. Later, 8.0% PAGE was conducted to separate SSR primers-amplified products for 2–2.5 h at 160 V, while rapid silver staining (Liu et al., 2008) was performed for visualization.

The AxyPrep DNA gel extraction kit (Axygen Bio Inc., Hangzhou, China) was utilized to recover part of amplified products with desirable size from PAGE gels. Meanwhile, T-A cloning kit (TaKaRa) was adopted for cloning those products extracted, whereas ABI 3730 (Applied Biosystems, United States) was adopted for sequencing the positive clones at Beijing Genomics Institute (BGI Shenzhen, China).

Survey of Polymorphism and Genetic Diversity Analysis

In order to further estimate the application of these SSR markers and validate the polymorphism of these loci, 32 diverse cabbage cultivars were selected and classified according to the predicted genetic distance (**Supplementary Table S1**). For SSR markers, we calculated their polymorphic information content (PIC) values by using Power Marker v. 3.0 (Liu and Muse, 2005). Genetic similarity coefficients across the diverse accessions were calculated using NTSYS-pc software SIMQUAL program using the 0–1 data matrix. Moreover, dendrograms were constructed by applying NTSYS-pc software SAHN module *via* the unweighted pair-group method with arithmetic averages (UPGMA; Rohlf, 2000; Kumar et al., 2001).

Furthermore, to facilitate the efficient use of primers and enable them to be easily operated, a strategy designated MCID was adopted, where cultivars were identified by the manual scoring and selection of certain bands (Wang et al., 2011; Korir et al., 2013; Zhai et al., 2014). We distinguished the 32 cabbage genotypes clearly according to other SSR markers used with certain band sizes. In addition, for assessing SSR markers for their amplification efficiency and transferability, we amplified 24 primer pairs of SSRs in 10 relevant crop species belonging to Brassicaceae family by adopting the above-mentioned PCR conditions.

RESULTS

The SSR Content of the Cabbage Genome and Cross-Species Comparison

Altogether 64,546 perfect SSR motifs (132.01 SSR/Mb) were identified from the 0.5 Gb in genomic sequence of cabbage, including 3,338 compound SSRs (**Table 1**). In addition, there were 93,724 imperfect SSR motifs (**Table 2**). Then, we compared SSRs distribution and level between the 'TO1000' cabbage genomic sequence and 11 additional genomes of related plant species to varying levels (the sequence was 5.5 Gb in length, about 0.8 million SSRs). Later, the related information was obtained based on the databases (**Supplementary Table S2**).

[2]http://pgrc.ipk-gatersleben.de/misa/primer3.html
[3]http://www-genome.wi.mit.edu/genome_software/other/primer3.html

TABLE 1 | A comparative survey of perfect Simple sequence repeats (SSRs) across 12 analyzed genome sequences.

Genome	Analyzed sequences (Mbp)	Perfect SSRs				Compound SSRs	
		Count	Density (SSRs/Mbp)	Cumulative (Mbp)	Cumulative (%)	Count	%
Brassica napus L.	976.191	123,212	126.22	2.70	0.28	7,297	5.92
Brassica rapa subsp. pekinensis	284.129	42,656	150.13	1.05	0.37	2,223	5.21
Arabidopsis thaliana (L.)	119.669	17,225	143.94	0.35	0.29	737	4.28
Brassica cretica	412.521	65,262	158.20	1.33	0.32	1972	3.02
Raphanus sativus L.	402.328	49,605	123.29	1.15	0.29	2,508	5.06
Brassica nigra	402.145	52,117	129.60	1.24	0.31	2,932	5.63
Brassica juncea var. tumida	954.861	104,035	108.95	2.22	0.23	9,970	9.58
Camelina sativa	641.356	135,740	211.65	3.06	0.48	8,845	6.52
Capsella bursa-pastoris	268.431	46,394	172.83	0.93	0.35	2,479	5.34
Eutrema yunnanense	415.364	53,260	128.22	1.10	0.26	3,337	6.27
Barbarea vulgaris	167.352	34,939	208.78	0.67	0.40	2,764	7.91
Brassica oleracea	488.954	64,546	132.01	1.40	0.29	3,338	5.17

The number of perfect SSRs found in the *B. oleracea* genome was similar to those of *B. cretica* (65,262), *E. yunnanense* (53,260), *B. nigra* (52,117), *R. sativus* (49,605), *C. bursa-pastoris* (46,394), and *B. rapa* (42,656). The *B. oleracea* genome was also found to contain almost four times as many perfect microsatellites as that of *A. thaliana* (17,225), and twice the number compared with *B. vulgaris* (34,939). However, it contained only half of those in *C. sativa* (135,740), *B. napus* (123,212), and *B. juncea* (104,035). The cumulative length of the full collection of cabbage SSRs was 1.4 Mbp, which comprises 0.29% of the assembled genome. The same percentage was found in *A. thaliana* (0.29%) and radish (0.29%) but considerably lower than that found in *C. sativa* and *B. vulgaris* (0.48% and 0.40, separately). Compound SSRs represented 5.17% of the cabbage perfect SSRs, which only exceeded those of *A. thaliana, B. cretica* and radish (**Table 1**).

Characterization of the SSR Motifs by Different Lengths and Repeats

The cabbage SSR motifs that predominated were the Di- and Mono- (40.9 and 29.2% of all the SSRs, respectively, with densities of 54.04 and 38.60 SSRs/Mbp, respectively), with smaller proportions of trinucleotides (25.06%) and tetranucleotides (6.87%); the penta- and hexanucleotide repeats contributed <5% (**Table 2**). Dinucleotide sequences played dominant roles, which constituted 0.64 Mbp (45.44% of accumulated length for total SSR motifs). Dinucleotides accounted for the most frequently seen type within tomato and eggplant. Of those imperfect SSR motifs, there were less mono- to tetranucleotide motifs than those seen in perfect SSRs group. Besides, there were larger motifs, which together with penta- to hexa-SSRs, accounted for 22.3% of the accumulated length for overall imperfect SSR motifs (**Table 2**). It could be discovered that the sum of Di- and Mono- formed the majority of perfect SSRs in all the genomes of Brassicaceae family that were searched and the species, including *C. sativa* and *E. yunnanense*. However, the majority of perfect SSRs in genomes of radish and *C. bursa-pastoris* are primarily formed by di- and trinucleotides. In the *A. thaliana* and *B. vulgaris* genomes, the sum of mono- and trinucleotides was the most frequent type (**Figure 1**).

Figure 2 and **Supplementary Table S4** present changes in perfect SSR motifs within cabbage genome in terms of repeat unit number. According to our results, larger repeat motifs had significant reduction compared with mono- and di-nucleotide types, among which, tetra- to hexa-nucleotide types experienced the most significant decrease as repeats increased (**Figure 2A**). Consequently, there were over double dinucleotide repeat units (11.8) relative to hexanucleotide (5.1) and trinucleotide (6.5) ones, and they were about thrice of penta (4.4) and tetra- (4.5) ones (**Table 2**). According to perfect repeat motif length, we considered 88.6, 9.8 and 1.6% of SSRs as hypervariable class I possibly variable class II and variable class III, respectively (<20, 20–30, ≥30 nt, separately; **Figure 2B**). All the types of nucleotides are members of class I (**Figures 2B,C**).

Characterization of SSRs by Classified Type

We classified repeats according to previous description (Jurka and Pethiyagoda, 1995). Therefore, the class AAT of trinucleotide repeats contained (ATA)n, (TTA)n, (TAT)n, (ATT)n, and (TAA)n, and they were the same in terms of diverse reading frames or complementarity. We discovered altogether 331 SSR motif types, and the potential base combinations included mono- ($n = 2$), di- ($n = 4$) tri- ($n = 10$), tetra-nucleotides ($n = 31$), together with 80 penta-nucleotide repeat variants and 204 hexanucleotide repeat variants (**Table 2**).

In this study, the individual repeat motifs for each type of

TABLE 2 | Variation in the repeat length among genomic cabbage perfect and imperfect SSRs.

SSR type	Perfect motif							Imperfect motif		
	Kinds	Count	%	Density (SSRs/Mbp)	Cumulative (Mbp)	Cumulative (%)	Mean repeat Number	Count	%	Density (SSRs/Mbp)
Mono-	2	18,876	29.24	38.60	0.36	25.79	19.1	19,916	21.25	40.73
Di-	4	26,425	40.94	54.04	0.64	45.44	11.8	31,612	33.73	64.65
Tri-	10	12,253	18.98	25.06	0.25	17.70	6.5	13,061	13.94	26.71
Tetra-	31	4,433	6.87	9.07	0.08	6.00	4.5	8,235	8.79	16.84
Penta-	80	1,505	2.33	3.08	0.03	2.48	4.4	13,952	14.89	28.53
Hexa-	204	1,054	1.63	2.16	0.03	2.44	5.1	6,948	7.41	14.21
Total/mean	331	64,546	100.00	132.01	1.40	100.00	12.1	93,724	100.00	191.68

SSR in the cabbage genome were also evaluated (**Figure 3** and **Supplementary Table S5**). The base composition of cabbage SSR motifs is strongly biased toward A and T. The most frequent mono- to hexa-nucleotides motifs were A/T (97.3%), AT/AT (67.9%), AAG/CTT (33.7%), AAAT/ATTT (42.4%), AACCG/CGGTT (23.4%) and AAAAAT/ATTTTT (8.5%). Regarding the distribution of different motifs, the AT repeats were not only the predominant dinucleotides, they were also the most frequent motif in the entire genome, comprising 32.3% of the total SSRs. Alternatively, CG repeats were barely detected. AAT, AAG, AAC, and ATC repeat types occupied the predominant roles in trinucleotide motifs (78.3% in total), while GC-abundant repeats, including CCG, AGC, and ACG, showed low abundances.

Consistently, AT-abundant tetranucleotide motifs, like AAAG, AAAT, AATT, and AAAC occupied the predominant role within cabbage genome (72% totally), whereas motifs AAAAT, AACCG, AAAAG, and AAAAC accounted for 57.0% of overall pentanucleotide repeats. There were just 4 hexanucleotide motif types, including AAAAAG, AAAAAT, AAGCCC, and AAAAAC existing, and the abundance was >5% (**Figure 3**). The close motif-type distribution was observed among nearly every remaining species detected in the present work (**Supplementary Table S6**).

The Distribution of SSRs in the Chromosomes

Those SSR loci within cabbage genome discovered were further classified according to the corresponding distribution and motifs in pseudomolecules. There were 7,172 perfect together with 10,414 imperfect SSRs in C1-C9 discovered from pseudomolecules on the whole (**Table 3** and **Supplementary Table S7**). We also predicted the association of chromosome length with SSR number, and a great correlation coefficient $R^2 = 0.9653$ was obtained (**Figure 4A**). There were most SSRs in C3 group (longest linkage; including 13,461 imperfect and 9,311 perfect, 64.98 Mbp), whereas C6 group (shortest linkage) had least SSRs (including 7,839 imperfect and 5,393 perfect, 39.82 Mbp). Nonetheless, there were great differences in SSR density across diverse chromosomes, which were between 129.55 (C2) and 143.29 (C3) perfect, whereas between 189.79 and 207.16 imperfect SSR/Mbp, separately (**Table 3**). It was observed that the distribution of motif types within individual chromosomes was very similar to the pattern found over the whole genome, with the mono- and di- repeats observed the most frequently and penta- and hexa- the least (**Figure 4B**). The number of SSR motifs on each chromosome (C1-C9) ranged from 5,393 (C6) to 9,311 (C3; **Figure 4C**). Mono- and di- SSRs exhibited maximum variation among linkage groups, with C1 and C5 exhibiting the lowest percentages for di- (38%) and mono- (28%), respectively, and the highest for tri-nucleotides (20%). When diverse motif distributions on the chromosome were considered, the commonly seen mono- to trinucleotides had close proportion to that acquired from the whole-genome. However, the relative contributions of the tetra-, penta- and hexanucleotides varied greatly between different linkage groups (**Supplementary Table S5**).

Gene Context of SSRs

The genomic distribution of SSRs was compared with their association with individual genes based on the data from assembled chromosomes of cabbage genome (Gadaleta et al., 2012; Bhattarai et al., 2021). A total of 14,349 perfect SSRs (22.23%) and 18,889 imperfect SSRs (20.15%) were associated with 11,013 (18.15%) and 13,707 (22.58%) genes, respectively (**Table 4**). It accounts for 18–22% of the entire gene space. It was predicted that such cabbage genes covered altogether 29.38 Mbp; in other words, the density was 138.24 for perfect whereas 181.97 SSRs/Mbp for imperfect motifs, separately, in the gene space. Based on repeat motifs, we divided SSRs distribution on every pseudomolecule (**Supplementary Table S8**). **Figure 5** presents the comparisons of SSR motifs discovered from genomic and genetic sets. We assigned the overall SSRs populations within the gene space and genome as non-triplet repeats (mono-, di-, tetra-, penta-nucleotides) and triplet repeats (tri-, hexanucleotides). As for imperfect (38.89%) and perfect (40.7%) motifs, there were more gene sequences in triplet repeats (**Figure 5A**). Typically, trinucleotides represented the most frequent type, occupying 38.58% (53.3 SSR/Mbp) for perfect whereas 31.04% (56.49 SSR/Mbp) for imperfect genic SSRs, separately (**Table 4** and **Figure 5B**). The most common dinucleotides were AT/AT. They comprised 21.1% of the total genic SSRs. The most frequent genic SSR motif types were the trinucleotides AAG/CTT (31.0%), ATC/GAT (17.1%), AGG/

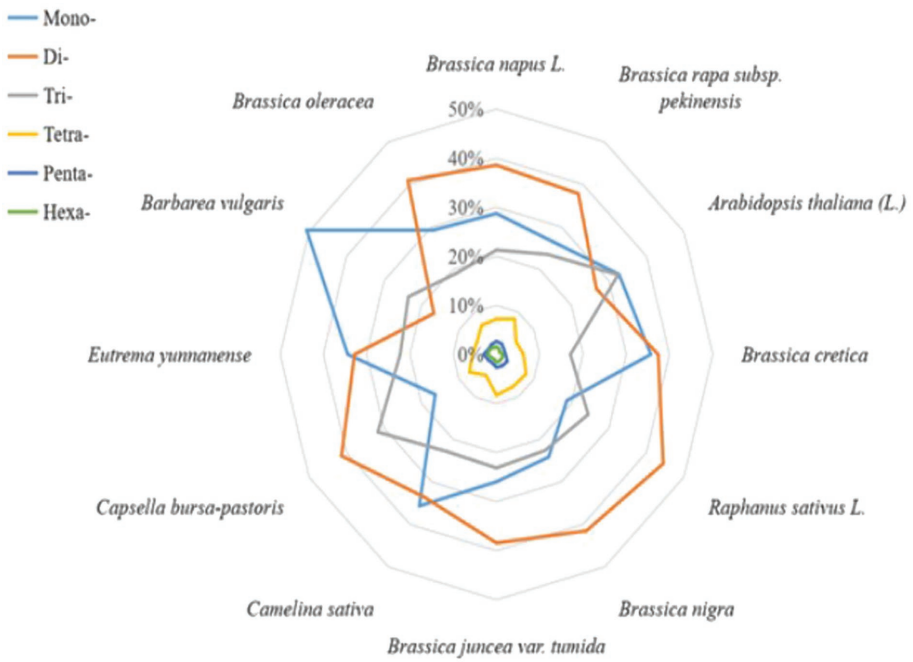


FIGURE 1 | Distribution of SSR repeated motif length across 12 cruciferous genomes.

CCT, and AAC/GTT (**Figure 5C**). Therefore, we compared a group of SSR genes in the cabbage reference gene space and evaluated the specific gene regulation functions that are frequently present. The genes that contained one or more SSRs were discovered within 60 sub-GO categories ("biological processes" (BP), "cellular components," (CC) and "molecular function" (MF); **Figure 6** and **Supplementary Table S9**). Over-representation was found for a number of gene families, such as BP in the sub-categories "Xylem and phloem pattern formation" (GO:0010051), "Potassium ion transmembrane transport" (GO:0071805), and "Regulation of gene expression"(GO:0010468); for MF, "Microtubule binding" (GO:0008017) and "Microtubule motor activity" (GO:0003777). No enrichment was observed for CC.

Development and Validation of Unique SSR Primer Pairs

We also obtained flanking sequences for each SSR motif within cabbage genome, which were adopted to be the targets to design primers. For obtaining specific primer pairs, they were aligned against reference genome of cabbage according to primer selection criteria. Finally, altogether 64,546 primer pairs were obtained (**Supplementary Table S10**).

This study prepared 1,113 SSR primers (**Supplementary Table S11**) at random, analyzed them for preliminary verification, and amplified them by 2 DNA templates of cabbage, namely, 'JSC142' and 'JSCJF'. Among the SSR primers, 916 pairs (82.3%) were stably and repeatedly amplified. For better confirming whether sequences that contained polymorphic microsatellites were real and positive, we recovered and sequenced 30 co-dominant segregation segments following T-A cloning. As a result, these sequences conformed to the initial ones, which indicated the high specificity of our prepared SSR primers.

Genetic Diversity Analysis of Cabbage Genotypes

This study prepared altogether 60 possible SSRs to conduct PCR validation by using the PAGE gels. Among them, 32 primers exhibited diacritical polymorphisms across diverse genotypes (**Figures 7A–C**). For investigating the possibility of using those candidate SSRs to carry out genetic analysis, we chose 32 SSR primers showing polymorphism for assessing genetic diversity for those 32 cabbage cultivars obtained from diverse areas (**Supplementary Table S1**). We discovered altogether 105 alleles, including 92 (87.67%) polymorphic alleles. On average, there was 2.9 alleles at each locus (range, 1–8). Additionally, the average PIC value was 0.46 (range, 0.14–0.88; **Table 5**). The sizes of the amplicons for the SSRs markers ranged from 134 bp to 273 bp. The information of these informative SSR primers is shown in **Table 5**.

The genotype data were analyzed *via* NTSYS-2.10e software, and the dendrogram showed that 32 cabbage accessions could be classified into three major clusters with similarity coefficients that ranged from 0.46 to 1.00 (**Figure 7D**). Clusters I and II included 17 and 11 accessions, respectively. Most of the accessions in these two clusters had different geographical origins, leaf colors, ball shapes, and maturities. In Cluster I, JSC28 presented a high similarity with JSC40. Both originated in China and have green leaves and mature early. JSC37 and JSC111 were divided into a subgroup. Both originated in Europe and are gray-green. JSC3 that originated in Japan and JSCJF that

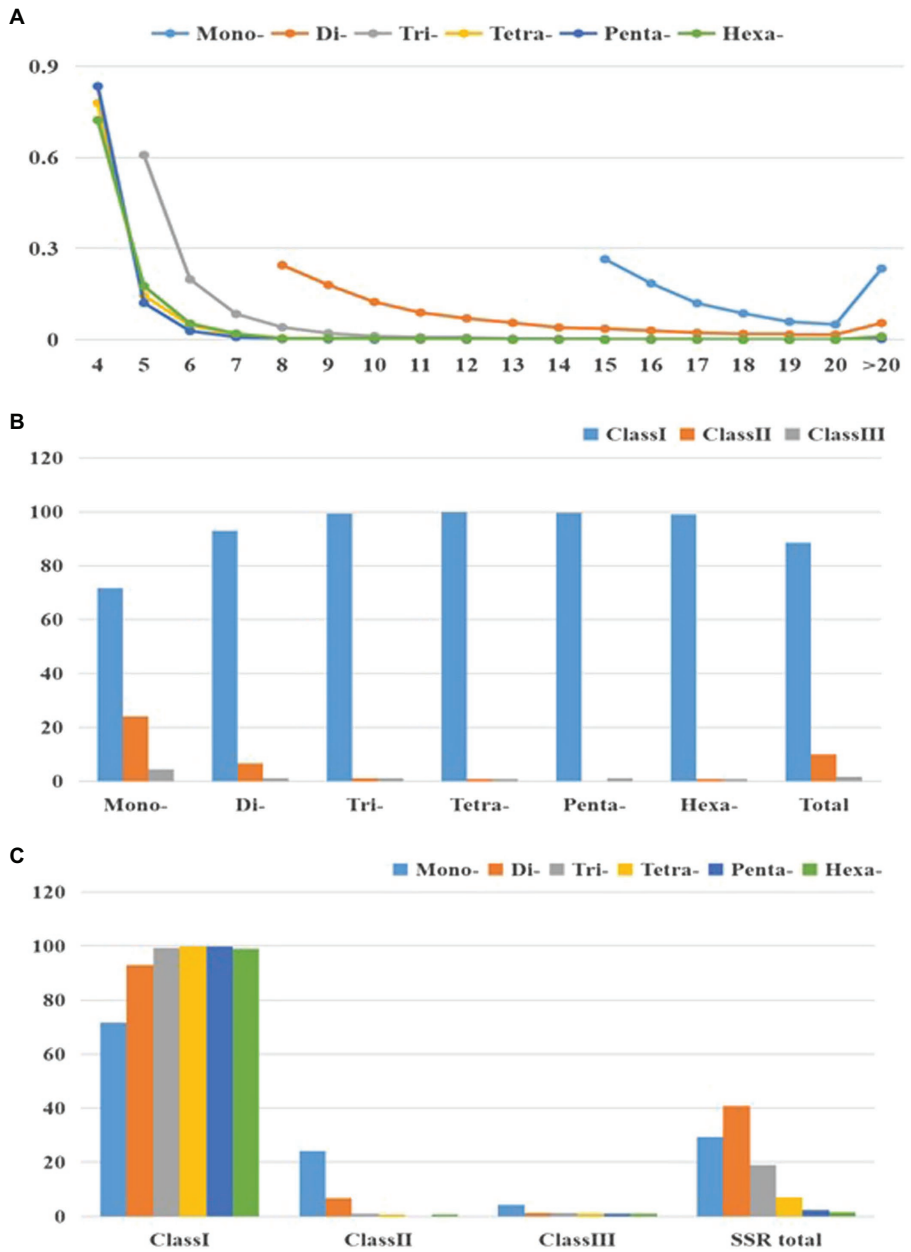


FIGURE 2 | Characterization of perfect SSRs in the cabbage genome. **(A)** The changeable rule from mononucleotides to hexanucleotide motifs. **(B)** The frequency of repeat classes (class I > 30 nt, class II 20–30 nt, class III <20 nt). **(C)** The distribution of motif type within each class.

originated in China were assigned to one subgroup, since both have a similar leaf color and ball shape. JSC13 that originated in China, and JSC90 that originated in Japan were also classified as one subgroup, because they had round ball morphology, green leaves, and early maturation. We classified JSC185 originating in Japan and JSC2 originating in China as one subgroup. Although both have a similar leaf color, they have different spherical characteristics and maturities.

In Cluster II, JSC10 that originated in China and JSC168 that originated in the Netherlands were assigned to one subgroup. Both have the same ball shape. They differ in that JSC10 has yellow-green leaves and matures extremely early, while JSC168 has purple leaves and matures late. JSC18 that originated in Japan and JSC43 that originated in China were assigned to one subgroup, and both have green leaves, a round ball shape and mature early. JSC23 and JSC30 were assigned to one subgroup. Despite that they differed in origin, they have the same spherical shape, leaf color and stage of maturity. A wild accession JSCYS, which was collected in the United States, and it has dark green leaves and a non-heading character was assigned to a separate subgroup in Cluster II. Cluster III comprised four accessions. All were collected from China and

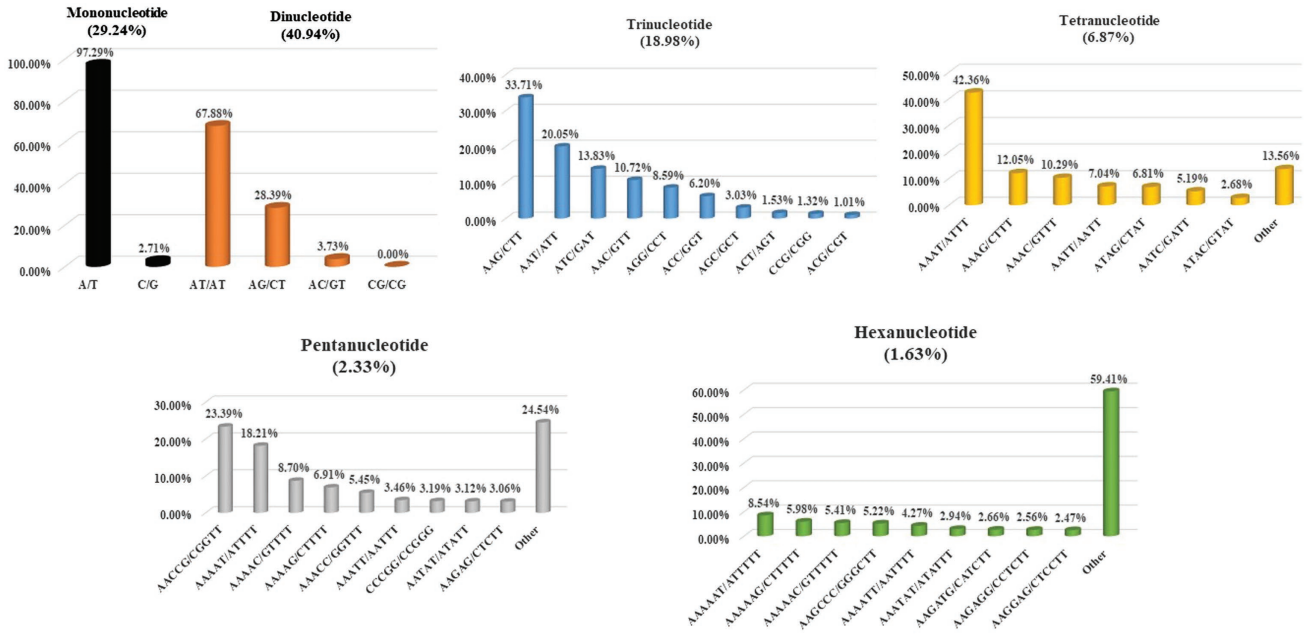


FIGURE 3 | Distribution of the major repeat motifs in cabbage genome.

TABLE 3 | The chromosome-by-chromosome distribution of perfect, compound, and imperfect SSRs.

Linkage groups	Total Mbp	Perfect								Imperfect	
		Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	Total	SSRs/Mbp	Total	SSRs/Mbp
C1	43.76	1,662	2,351	1,158	397	114	93	5,775	131.97	8,418	192.37
C2	52.89	1944	2,949	1,211	473	164	111	6,852	129.55	10,038	189.79
C3	64.98	2,792	3,832	1786	577	205	119	9,311	143.29	13,461	207.16
C4	53.72	2075	3,008	1,371	503	188	112	7,257	135.09	10,322	192.14
C5	46.9	2007	2,414	1,207	426	130	115	6,299	134.31	8,944	190.70
C6	39.82	1,525	2,190	1,069	393	122	94	5,393	135.43	7,839	196.86
C7	48.37	1846	2,555	1,240	469	160	133	6,403	132.38	9,501	196.42
C8	41.76	1735	2,349	1,132	407	162	107	5,892	141.09	8,512	203.83
C9	54.68	2,127	3,050	1,424	512	177	111	7,401	135.35	10,683	195.37
Total	488.954	18,876	26,425	12,253	4,433	1,505	1,054	64,546	132.01	93,724	191.68

have the same spherical shape. JSC7 and JSC410 have the same color leaves and state of maturity. JSC12 has gray-green leaves and matures extremely late, while JSC107 is blue-green and matures at a medium stage.

MCID of Cultivar Identification With SSR Markers

This study identified 32 cabbage cultivars by using 5 SSR primers that contain the polymorphic and reproducible bands. Of those 5 primers utilized, the Bol-SSR32 primer was initially selected to identify cabbage genotypes (**Figure 8**). Based on PAGE analysis, the Bol-SSR32 primer produced 2 polymorphic bands within those 32 cultivars (**Figure 7A**), and it might be used to classify diverse cabbage genotypes to 3 groups according to with/without the characteristic 280 and 380 bp bands. Later, the SSR7 primer was used to separate those cultivars in 3 groups singly or to smaller groups, like JSC142 or JSCYS. Afterwards, we used the rest 3 primers to distinguish cabbage cultivars step by step. Typically, applying the Bol-SSR23 primer helped to separate those 32 cultivars from MCID (**Figure 8**). We screened several clear polymorphic primers during the course of experiment. It should be highlighted that only the clear polymorphic bands amplified with each primer were accurately used to differentiate the accessions. These indicated that the MCID method used in this study is a valuable and efficient strategy for the identification of cultivars in cabbage.

Application of SSR Primers to Other Species in the Brassicaceae

This study randomly chosen 24 stable and reliable SSR primers for amplification on 10 different species in the Brassicaceae family to identify the potential transferability and availability (**Figure 9**). In total, 21 of the 24 (87.5%) SSR primers exhibited transferability and applicability to one or more of the 10 related

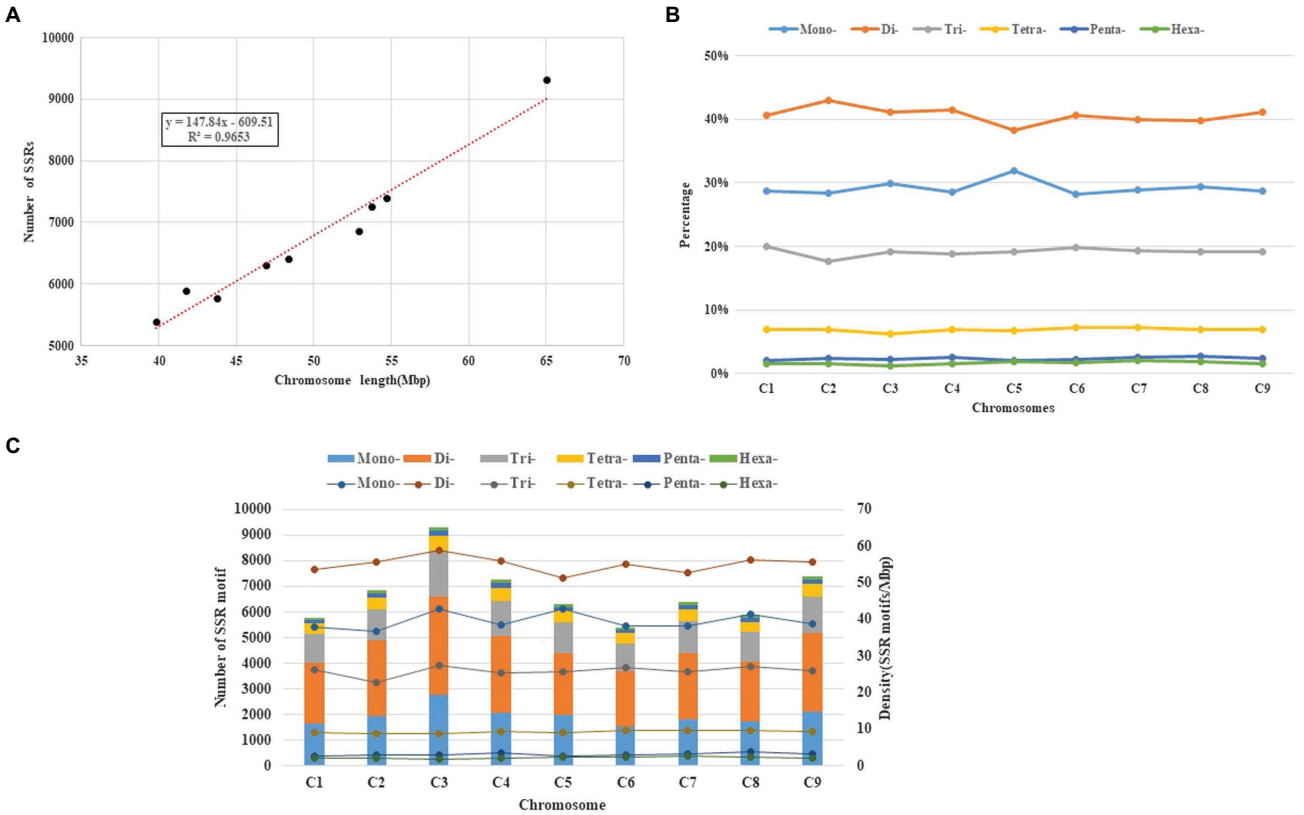


FIGURE 4 | Intra-chromosomal distribution of SSRs. **(A)** Relationship between the SSR number and chromosome length in the cabbage genome. **(B)** The frequency of mono- to hexanucleotide motifs in cabbage chromosomes. **(C)** Number and density of the SSR motifs across nine chromosomes of the cabbage reference genome.

TABLE 4 | Variation in repeat length among genic cabbage perfect and imperfect SSRs.

SSR type	Perfect motif			Imperfect motif		
	Count	%	Density(SSRs/Mbp)	Count	%	Density(SSRs/Mbp)
Mono-	3,632	25.31	34.99	3,768	19.95	36.30
Di-	3,697	25.76	35.62	4,456	23.59	42.93
Tri-	5,536	38.58	53.33	5,864	31.04	56.49
Tetra-	951	6.63	9.16	1,535	8.13	14.79
Penta-	229	1.60	2.21	1785	9.45	17.20
Hexa-	304	2.12	2.93	1,481	7.84	14.27
Total/mean	14,349	100.00	138.24	18,889	100.00	181.97

Brassica species that were used in this study. Of them, altogether 9 primers exhibited different and stable bands among those 10 species, demonstrating the reliability and applicability of our identified SSR markers for cabbage in certain relevant Brassicaceae family members.

DISCUSSION

SSR markers have been deemed as the promising candidates to conduct genetic mapping and diversity analyses on crop species because they are specific and highly conserved (Yang et al., 2015). More and more articles have revealed that it is a highly effective and low-cost way to identify SSR markers on the basis of NGS. We collected and identified the genome-wide data of several cruciferous crops, including *B. oleracea,* to develop more SSR markers. A total of 64,546 perfect and 93,724 imperfect SSR were identified. Functional markers were employed in the genetic diversity analysis, to identify the cultivars among 32 different genotypes of cabbage and in the availability analysis across 10 relatives in the cruciferous crops.

The distribution of SSRs was examined within 12 genomes in the Brassicaceae family. Besides, it was discovered that genome size showed positive correlation with SSR motif number

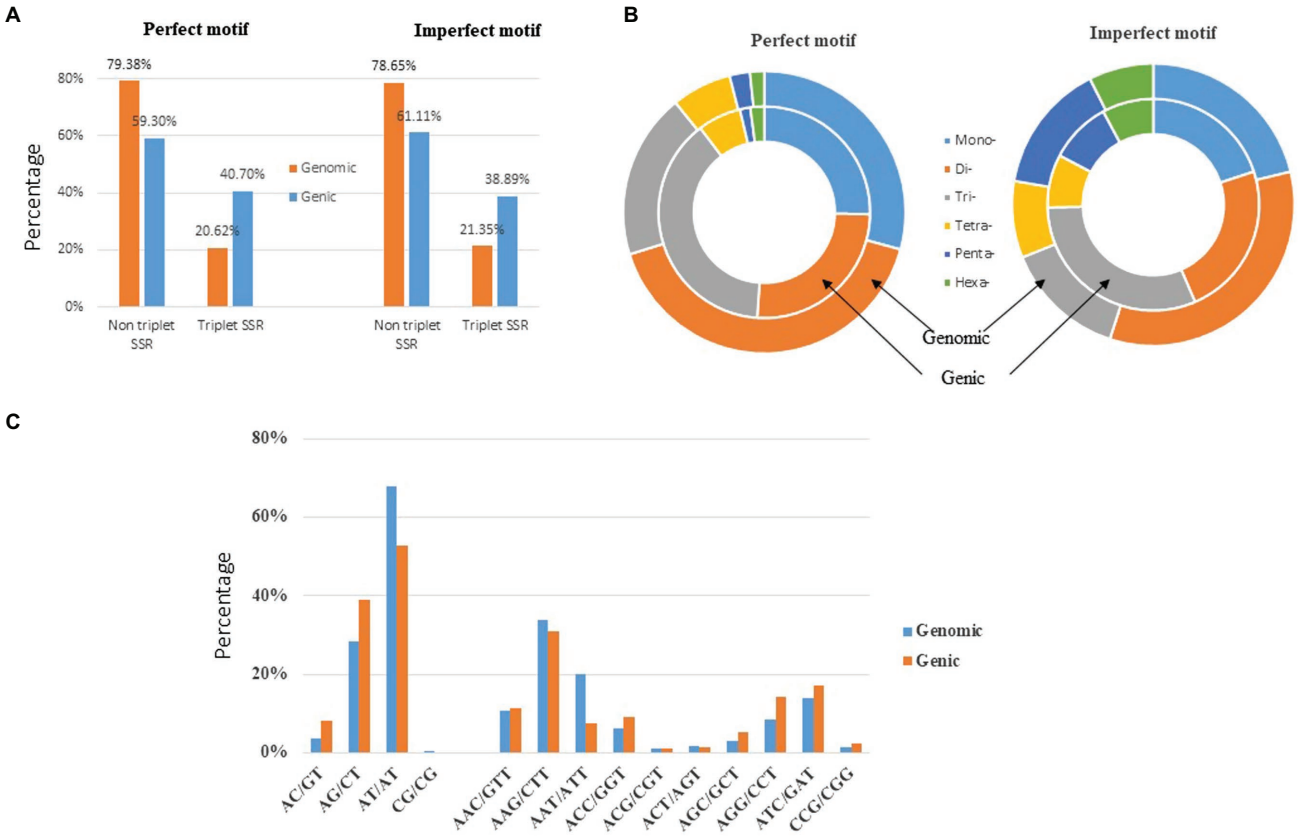


FIGURE 5 | Distribution of microsatellite sizes in the cabbage genome. **(A)** Non-triplet SSR vs. triplet SSR in both perfect and imperfect motifs. **(B)** Distribution of repeat types within perfect and imperfect SSR motifs. **(C)** A comparison between di- and trinucleotide repeats in both the gene space and full genomic region.

discovered after comparing the 12 species. Several reports showed that species that possessed larger genomes typically display a lower SSR density (SSRs/Mb; Morgante et al., 2002; Portis et al., 2018). Nonetheless, the different genomic sizes will result in different microsatellite repetition levels, but SSR density is not associated with genome size (Behura and Severson, 2014; Portis et al., 2016). In this study, the three species *B. rapa*, *C. bursa-pastoris* and *B. vulgaris* were found to have larger genomes and exhibit a lower density of SSR. However, the species *B. nigra* and *E. yunnanense* are exceptions. Their genomes sizes are 402.1 and 415.4 Mbp, but their microsatellite densities are comparable to those found in *B. napus* that has a genome twice as large. The density of perfect microsatellites in the *B. oleracea* genome is the fourth highest observed within the *Brassica* family, even though it is similar to those detected in *B. nigra* and *E. yunnanense*. Consequently, SSRs are highly enriched and abundant within Brassicaceae, making them the attractive molecular markers to carry out genetic analyses of Brassicaceae (Lv et al., 2017).

In this study, we compared the classified types of SSR motifs among all the species in the Brassicaceae family. Variation in the selective constraint on sequence repeats could differ among the SSR motifs. Overall, cabbage had the second greatest number of motif types compared with the other members in this family. It is still unknown about whether such heterogeneity is associated with the species genomic evolution or complexity.

Dinucleotides were considered as the most common type in eggplant and tomato. In our study, although dinucleotides are the most common repeats in the cabbage genome, trinucleotides prevail in the gene space, which is analogous to those of other species (Cavagnaro et al., 2010). In contrast, tri- and hexanucleotides are the most common type in the gene space of eggplant genome. The reason could be attributed to negative selection against frameshift mutations in the coding regions, and because of the mutation pressure combined with possible and positive selection for specific single amino acid stretches, the trinucleotides have enhanced their frequency in the coding portion. Generally, AT-rich motifs occur more often in dicotyledons. It has been reported that AT-rich repeats are widespread in dicotyledonous but not in monocotyledonous species, and the difference between them may be partially accounted for by the nucleotide composition of their genomes. The monocots have a GC content of 43.7% compared with one of 34.6% in the dicots (Portis et al., 2018). We found similar results in the cruciferous crop genomes. However, the classified motif types were not completely identical within each species, which was shown by comparisons among several genomes in this study. The most common dinucleotides in

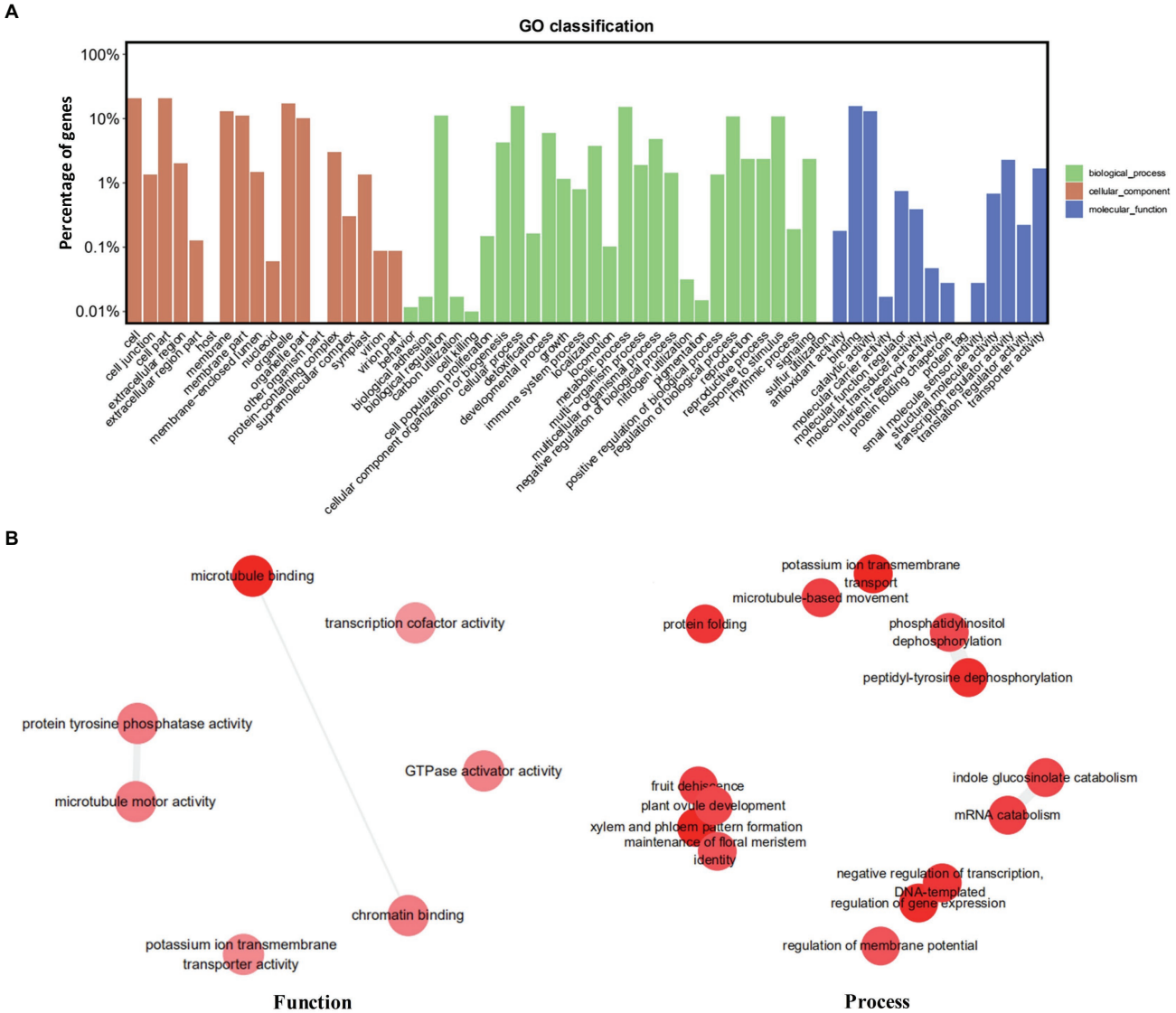


FIGURE 6 | Functional analysis (gene ontology) of cabbage genes containing SSRs. **(A)** GO classification. **(B)** Revigo summary of "biological process" and "molecular function" enriched terms.

this study were AT/AT, and the most frequent genic SSR motif types were the trinucleotides AAG/CTT. Similar patterns of motif distribution have been found in other species. For example, several studies reported that AT/AT repeats appear to be typical in non-transcribed regions, and AG/CT prevail in gene sequences, while AC/GT and CG/GC repeats are the least frequent dinucleotides in both genomic and gene sequences (Cavagnaro et al., 2010; Portis et al., 2016). GC-rich motifs showed a strong bias in their distribution in cabbage gene sequences, most notably for mono-, di- and tri. For example, a GC content of only 14% was discovered in the whole genomic trimeric SSRs, whereas the trinucleotide SSRs in genes had a GC content of 43%. Genic SSR markers show a higher efficiency among various species when compared with non-coding microsatellites, promoting their application as anchor markers suitable for comparative genetics analysis (Varshney et al., 2005). On the contrary, coding SSRs experience an increased selection pressure; as a result, only insufficient polymorphism degree can be provided for distinguishing the varieties with close relations. Nonetheless, the genetic SSRs of cabbage can offer a decreased number of possibly variable SSRs, because of the decreased corresponding repeat number compared with that within the whole genome; typically, among the SSRs, 63.04% contained ≤10 repeats, while just 8.43% contained ≥20 repeats. Over-representation was found in several subcategories, such as BP and MF, but no enrichment was observed. In previous studies, SSRs are reported to occur in certain gene functions within eggplants, globe artichoke and *Medicago truncatula,* while transcription factors (TFs) constitute a distinct gene class containing SSRs (Portis et al., 2016, 2018; Min et al., 2017).

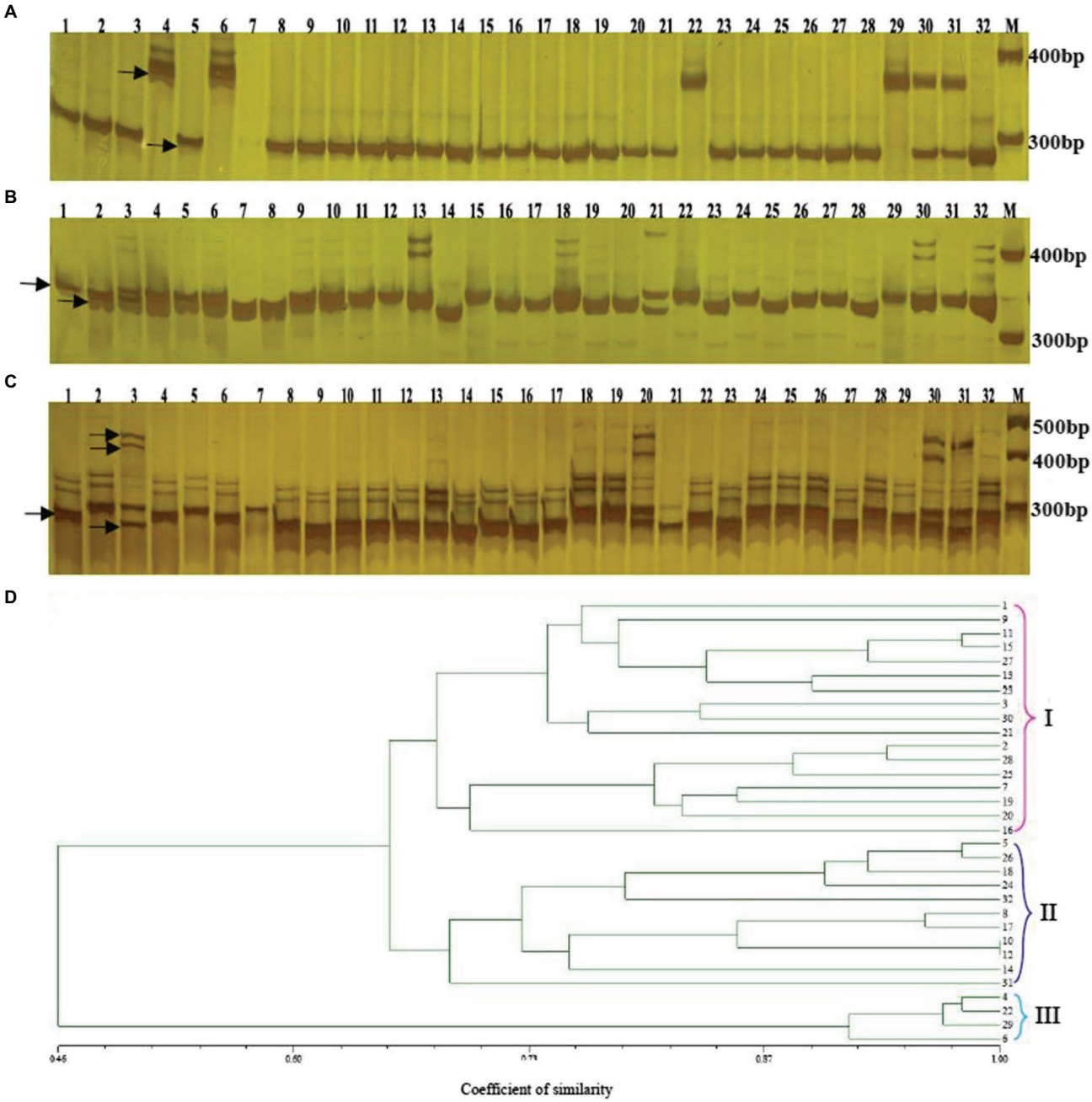


FIGURE 7 | Genetic diversity analysis of 32 cabbage accessions with SSR markers. Amplification of 32 genotypes with the SSR primers Bol-SSR32 **(A)** Bol-SSR24 **(B)** and Bol-SSR12 **(C)** by polyacrylamide gel electrophoresis, and the UPGMA dendrogram of 32 cabbage genotypes based on 32 new SSR markers **(D)** M: 50 bp DNA ladder. The genotype name of numbers (1–32) is listed in **Supplementary Table S1**.

In addition, TFs carrying SSRs have also been suggested to have critical functions, and their association with species diversity in Brassicaceae family should be clarified.

Longer repeats have been reported to show a lower abundance level within each class. In certain species, SSR frequency decreases as the repeat number increases, like globe artichoke and *Capsicum* (Cheng et al., 2016; Portis et al., 2016). For instance, SSRs that contain ≤10 repeats take up approximately 50% of the whole SSR number, whereas SSRs that contain >20 repeats only occupy <10%. According to our results, longer repeat motifs had significantly greater decreasing amplitude than mono- and di-nucleotide ones, among which, tetra- and hexa-nucleotide motifs exhibited the greatest decline in their frequencies as the repeat number increased. Some microsatellites were often found between neighbor genes that were reported to possibly be involved in gene regulation (Gao et al., 2013;

TABLE 5 | Genetic diversity analysis with SSR markers developed in cabbage.

Primer name	SSRID	Forward primer (5′-3′)	Reverse primer(5′-3′)	Expected size (bp)	Tm	Allele no.	PIC
Bol-SSR1	NC_027748.1:3927:SSR1	GTCGTCGTCGGTCATTAGGT	TATCCATCACCCCCTAACCA	200	60	2	0.14
Bol-SSR2	NC_027748.1:4466:SSR2	TCCTCATGGTCCCTTGCTAT	TGATACCTACCCAAGCAGGC	252	60	4	0.31
Bol- SSR3	NC_027748.1:7081:SSR3	GAGACTGAGACAGATCCCGC	TGGCGATGGTAGAGAAGAAAA	237	59	1	0.88
Bol- SSR4	NC_027748.1:8162:SSR4	CTCAGACGGTTGCAGATTCA	TCTCCGTTAACATGGCCTTC	169	60	2	0.24
Bol- SSR5	NC_027748.1:8546:SSR5	TACTAGGCGATGGCAGACCT	TCGTTGAATCCCACCTCTTC	134	60	2	0.16
Bol- SSR6	NC_027748.1:27390:SSR6	ATAAATGATGCGGCCAGAAG	GCAAGGAAGATCCATCGAAA	187	60	7	0.35
Bol- SSR7	NC_027748.1:31848:SSR7	CGAATTCGCTGTCATCTTCA	AGCCACCCGGTGAAACTTTA	229	61	5	0.21
Bol- SSR8	NC_027748.1:38054:SSR8	TTTCCGTAATCGCAAGGAAC	ATGGTCACCGACCTCAACAT	262	60	2	0.22
Bol- SSR9	NC_027748.1:38711:SSR9	GAAAAGCAGCACAACAACGA	CCACAAAACGTGTGCTTGAG	272	60	2	0.45
Bol- SSR10	NC_027748.1:48258:SSR10	CCCTCTCCTTTTTCTTTCGC	ATCTTTGCTGCTTCGGTCTC	238	60	1	0.51
Bol- SSR11	NC_027748.1:50103:SSR11	GTAGGGGTGGTCCAAAAGGT	CTTGGATCCTTCTCCCACAA	197	60	1	0.39
Bol- SSR12	NC_027748.1:51423:SSR12	GCTTTATACGCCGTCTCTCG	TTGAAAAGAGCCGCTCCTAA	186	60	8	0.47
Bol- SSR13	NC_027748.1:56608:SSR13	GTGTTTGAAGGAGAGCCGAG	GAACTAAAAAGAGCGCGTGG	197	60	3	0.42
Bol- SSR14	NC_027748.1:65043:SSR14	GGGAGATTATGGGCCTTCTC	TTGGTGGGATGGATGTGAAT	140	60	2	0.52
Bol- SSR15	NC_027748.1:65170:SSR15	GTACCAAAACGAAAGGCCAC	TGAAGTATCCAGAGGGGCAC	273	60	2	0.63
Bol- SSR16	NC_027748.1:116242:SSR21	CTTTTTCAGTCCGTAAGGCG	GCGGGTTACATCCGAAATTA	252	59	2	0.77
Bol- SSR17	NC_027748.1:104629:SSR17	TGAATGGCAAATTCCACAAA	ACGTTGAGATGGCAGGAATC	251	60	3	0.38
Bol- SSR18	NC_027748.1:105374:SSR18	TTATAGCAATCCCCACAGCC	CCATTGTCCTGGCTTTGATT	211	59	4	0.36
Bol- SSR19	NC_027748.1:105555:SSR19	AATCAAAGCCAGGACAATGG	CTCCTCAAGAGCACACTCCC	190	59	6	0.49
Bol- SSR20	NC_027748.1:108306:SSR20	TCGGCGGTTTCTTTTATTTG	CGTTTGTATTGAACCCAGCA	261	60	2	0.69
Bol- SSR21	NC_027748.1:116613:SSR22	GGCAATTGCACTAAATGACCA	TCTCGAGCTTCCCATCTTTG	256	60	2	0.65
Bol- SSR22	NC_027748.1:117330:SSR24	GATTCTAGTCCGGCGATGAC	AACTGGCCTTAATGGCTTCA	146	59	3	0.41
Bol- SSR23	NC_027748.1:135117:SSR25	GGCAAGGATGGACATGATCT	TTCAGAGGATGGAGAGCGAT	237	59	2	0.33
Bol- SSR24	NC_027748.1:293746:SSR58	CGCTGGACCACTTGTACTGA	CCGGCTAATTTACAGCTCCA	170	60	3	0.48
Bol- SSR25	NC_027748.1:135343:SSR26	TCGCTCTCCATCCTCTGAAT	TTGATTGATTGCTTGCTTGC	185	59	2	0.52
Bol- SSR26	NC_027748.1:137015:SSR29	CTCATCGTCGGGATCATCTT	GAGCAGAGTAGCGGAACCAC	192	60	5	0.23
Bol- SSR27	NC_027748.1:138313:SSR30	ATCCCTCCCCATTTTACCAG	GAGATGGCTAAGCGTCAAGG	134	60	4	0.19
Bol- SSR28	NC_027748.1:142981:SSR35	CTGAGACCAACGTGAGCGT	AAATTGGAGACGAAGGCAGA	231	60	2	0.81
Bol- SSR29	NC_027748.1:154301:SSR36	CACGTGAATTGCTCGTGTTT	GAACGTCGACGAATTTGGTT	226	59	2	0.78
Bol- SSR30	NC_027748.1:155819:SSR37	TCCGAAACTACCCCTCTCCT	CGATTACCTCCTGAAATCCG	155	60	3	0.67
Bol- SSR31	NC_027748.1:161172:SSR38	AAGCCAACCAACTCCCTGTA	TGGATCGATCTAAGCGAAAAA	251	59	1	0.32
Bol- SSR32	NC_027748.1:211682:SSR47	CCATTCGCAGCTGTATTTCA	ACCCACTGATGCATACCTCC	182	59	2	0.74

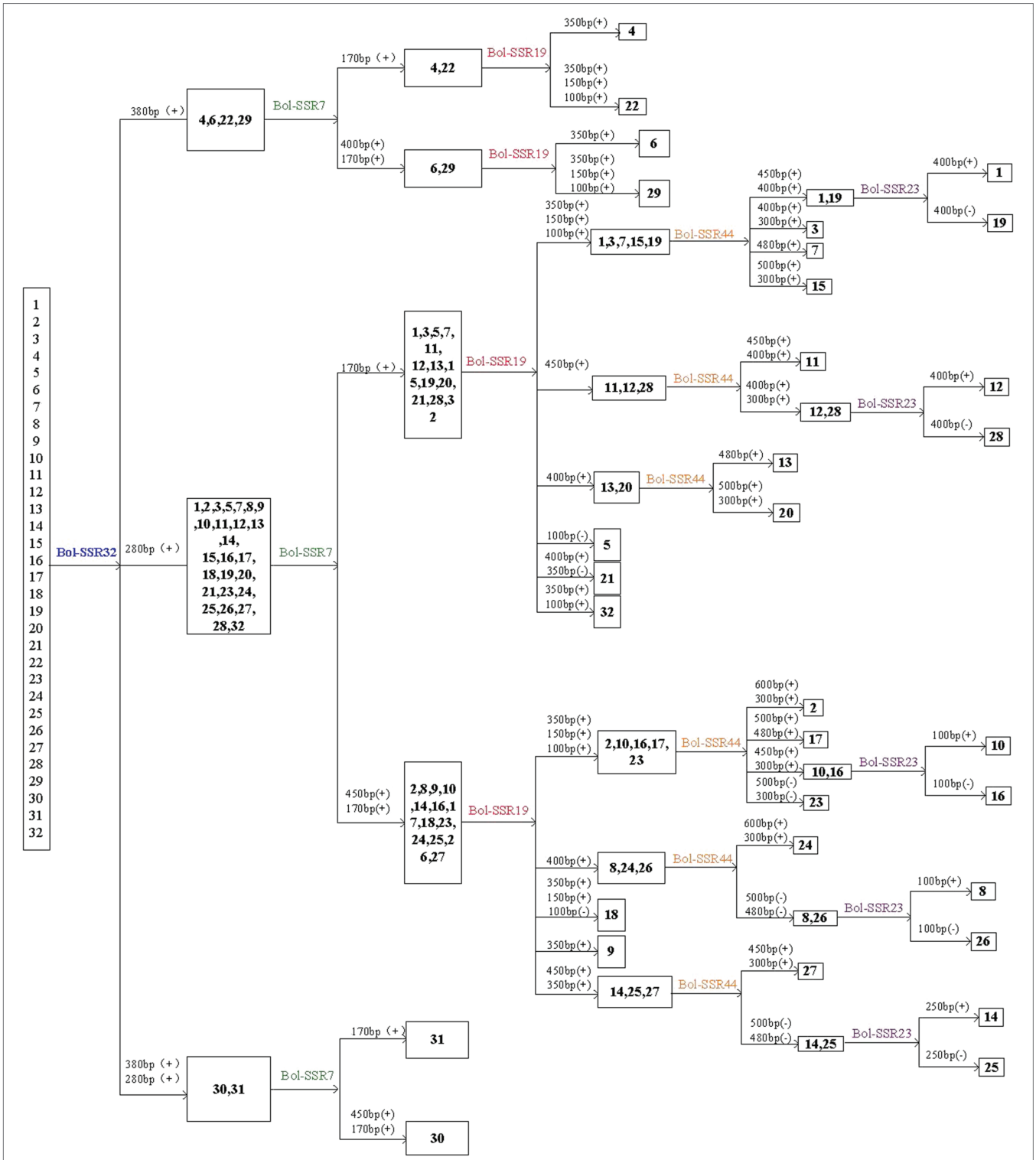


FIGURE 8 | MCID analysis of the 32 cabbage genotypes with the DNA fingerprints of five SSR primers. The number above each horizontal line in the diagram denotes the size of the polymorphic bands used to separate the genotypes following the line. (+) or (–) denotes the presence or absence of the polymorphic band.

Sawaya et al., 2013). Such microsatellite hotpots were also observed in this study, although they were mostly owing to long stretches of compound microsatellites. However, since most of the highly mutable loci are compound microsatellites that are composed of two or more repeated motifs, they could be exploited as putative highly polymorphic markers.

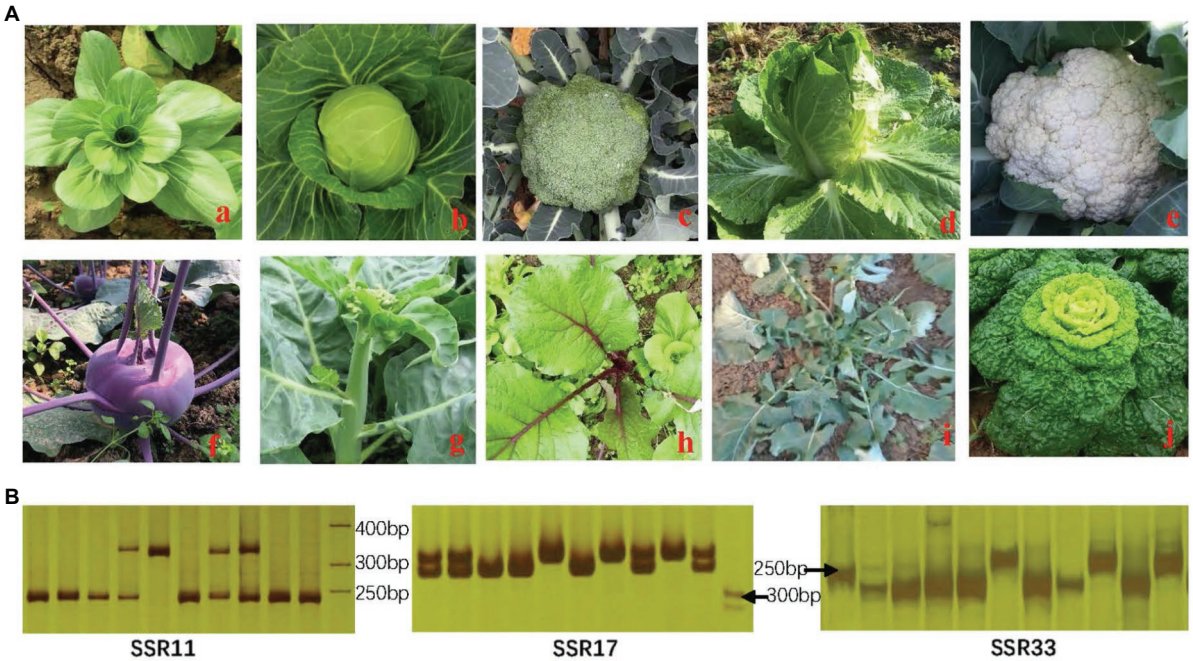


FIGURE 9 | Identification of related species in the Brassicaceae family **(A)** and amplified results by SSR primers **(B)**. (a) *Brassica campestris* subsp. *chinensis* (AA); (b) *Brassica oleracea* var. *capitata* (CC); (c) *B. oleracea* var. *italica* (CC); (d) *Brassica rapa* ssp. *pekinensis* (AA); (e) *B. oleracea* var. *botrytis* (CC); (f) *B. oleracea* var. *caulorapa*; (g) *B. alboglabra*; (h) *B. parachinensis*; (i) *B. napus* (AACC); (j) *B. campestris* L. ssp. *chinensis*.

However, genic SSRs have been demonstrated to exert an important role in gene expression and function in both humans and plants, which stand for a class of 'functional markers' in transcripts. They are also known as microsatellite instability (MSI), and MSI is known to enhance with plant development in *A. thaliana* (Golubov et al., 2010; Nelson et al., 2013). In previous studies, the occurrence of SSRs within specific gene functions has been found, and transcription factors form a significant class of genes that contain SSRs. Furthermore, the important role of transcription factors that possess microsatellites has been pointed out, and the relationship between this tendency and the species diversity of the Brassicaceae merits further study (Li et al., 2004).

To date, functional genetic markers, including SSRs, have progressively become a powerful approach to obtain insight into genetic studies owing to their multi-allelic detection, reproducibility and high cross-species transferability (Thiel et al., 2003; Taheri et al., 2018). With the emergence of NGS technology, the large-scale development of SSR markers based on genome-wide analysis directly or indirectly promotes the rapid development of marker-assisted breeding. A substantial number of SSR markers have been widely recognized in a variety of plants, including black pepper, pepper, pear, bitter gourd, bread wheat, *Camellia sinensis* and eggplant, by the analysis of genome-wide sequence data generated (Cui et al., 2017; Liu et al., 2018; Portis et al., 2018; Xue et al., 2018; Kumari et al., 2019; Uncu, 2019; Ahmed et al., 2020). In this study, a large number of SSR primer sets were the first ones to be comprehensively and successfully designed from the whole genome of cabbage, which is specific to previous studies. Many primers are able to amplify certain bands, whereas some can amplify the significantly greater bands, possibly because that the repeat number is changed or there is one small intron between primer pairs. Moreover, non-PCR fragment-producing primers might be associated with the existing huge introns or null alleles or the primer pair design among the splice sites.

Some recent articles have reported that SSR markers have been applied in the diversity and fingerprinting analyses within some plant species. According to our results, the whole-genome-based SSR markers showed high efficiency in distinguishing 32 cabbage species, and their distributions were not totally decided by the corresponding geographical sources, conforming to our prior works. According to the obtained results, the SSR markers extracted from the genome data of *B. oleracea* L. were suitable and served as excellent markers to distinguish cultivated landraces from wild ones. MCID is a new strategy that is more practical, economical, and effective at identifying plant cultivars with fewer primers, and the proposed method creates a recordable and readable flow chart, enabling the much easier identification of cabbage cultivars.

In addition, the genome-based genetic markers produced in this study are highly conservative and transferable from cabbage to some related cruciferous species, which is consistent with the results of research on cereals and the Leguminosae, Cucurbitaceae, and Rosaceae. However, the novel SSR markers developed with a relatively high level of transferability and availability will be conducive to advancing the investigation of comparative mapping analyses in the Brassicaceae family.

In brief, the SSR markers developed based on the WGS data in this study have polymorphism, repeatability, and transferability and will become an important tool for genetic mapping, germplasm identification and genetic diversity analysis of cabbage and its related species in the future.

CONCLUSION

In this study, a large number of potentially variable SSRs have been identified in cabbage. We identified 64,546 perfect SSR motifs and 93,724 imperfect SSR motifs in the 0.5 Gb of the cabbage genomic sequence, which was mined using a whole-genome bioinformatics survey. The cumulative length of full collection of cabbage SSRs was 1.4 Mbp, which comprises 0.29% of the assembled genome. Considering all Brassicaceae family members, the genome size was found to be positively associated with the number of SSR motifs identified. Dinucleotide sequences were the most common type in all cabbage SSR motifs, comprising 0.64 Mbp. As expected, microsatellites are ubiquitously distributed, and we detected a higher content of SSR repeats for longer chromosomes, as well as the homogeneous distribution of SSRs. Such innate characteristics of SSRs render them the suitable markers. Additionally, those 32 informative SSR markers chosen were adopted for determining the 32 cabbage genomes for their genetic diversity. Cultivars were efficiently identified by using the new strategy designating the manual diagram for identifying cultivars, and 5 Bol-SSR markers were utilized to distinguish 32 cabbage accessions. In addition, we also verified the transferability and availability of such SSRs based on additional 10 species belonging to Brassicaceae family. These results suggest that the genomic SSR markers that have been developed have considerable potential value in advancing cabbage research, including genetic mapping, MAS, and comparative genome analyses.

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 performed the data analysis and drafted the manuscript. MX, AZ, and LS conducted the validation of experiments and data analysis. WL and AZ contributed powerful analytical tools. AZ and JY helped with the revise of the manuscript. YX and AZ conceived and designed the research. All authors read and approved the final manuscript.

FUNDING

The study was supported by the Natural Science Foundation of Jiangsu Province (No. BK20190262).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.726084/full#supplementary-material

REFERENCES

Ahmed, H. G. M. D., Kashif, M., Rashid, M. A. R., Sajjad, M., and Zeng, Y. W. (2020). Genome wide diversity in bread wheat evaluated by SSR markers. *Int. J. Agric. Biol.* 24, 263–272. doi: 10.17957/IJAB/15.1433

Behura, S. K., and Severson, D. W. (2014). Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Res.* 22, 29–38. doi: 10.1093/dnares/dsu036

Bhattarai, G., Shi, A., Kandel, D. R., Solis-Gracia, N., da Silva, J. A., and Avila, C. A. (2021). Genome-wide simple sequence repeats (SSR) markers discovered from whole-genome sequence comparisons of multiple spinach accessions. *Sci. Rep.* 11:9999. doi: 10.1038/s41598-021-89472-0

Cai, X., Wu, J., Liang, J., Lin, R., Zhang, K., Cheng, F., et al. (2020). Improved Brassica oleracea JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. *Theor. Appl. Genet.* 133, 3187–3199. doi: 10.1007/s00122-020-03664-3

Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (Cucumis sativus L.). *BMC Genomics* 11:569. doi: 10.1186/1471-2164-11-569

Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., and Trejo-Saavedra, D. L. (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in capsicum. *Sci. Rep.* 6:18919. doi: 10.1038/srep18919

Cui, J., Cheng, J., Nong, D., Peng, J., Hu, Y., He, W., et al. (2017). Genome-wide analysis of simple sequence repeats in bitter gourd (Momordica charantia). *Front. Plant Sci.* 8:1103. doi: 10.3389/fpls.2017.01103

Gadaleta, A., Giancaspro, A., Giove, S. L., Zacheo, S., Incerti, O., Simeone, R., et al. (2012). Development of a deletion and genetic linkage map for the 5A and 5B chromosomes of wheat (Triticum aestivum). *Genome* 55, 417–427. doi: 10.1139/G2012-028

Gao, C., Ren, X., Mason, A. S., Li, J., Wang, W., Xiao, M., et al. (2013). Revisiting an important component of plant genomes: microsatellites. *Funct. Plant Biol.* 40, 645–661. doi: 10.1071/FP12325

Gil, J., Um, Y., Kim, S., Kim, O. T., Koo, S. C., Reddy, C. S., et al. (2017). Development of genome-wide SSR markers from angelica gigas nakai using next generation sequencing. *Genes* 8:238. doi: 10.3390/genes8100238

Golubov, A., Yao, Y., Maheshwari, P., Bilichak, A., Boyko, A., Belzile, F., et al. (2010). Microsatellite instability in Arabidopsis increases with plant development. *Plant Physiol.* 154, 1415–1427. doi: 10.1104/pp.110.162933

Ikten, H., Solak, S. S., Gulsen, O., Mutlu, N., and Ikten, C. (2019). Construction of genetic linkage map for Ficus carica L. based on AFLP, SSR, and SRAP markers. *Horticulture Environ. Biotechnol.* 60, 701–709. doi: 10.1007/s13580-019-00162-4

Jurka, J., and Pethiyagoda, C. (1995). Simple repetitive DNA sequences from primates: compilation and analysis. *J. Mol. Evol.* 40, 120–126. doi: 10.1007/BF00167107

Karci, H., Paizila, A., Topcu, H., Ilikcioglu, E., and Kafkas, S. (2020). Transcriptome sequencing and development of novel genic SSR markers from Pistacia vera L. *Front. Genet.* 11:1021. doi: 10.3389/fgene.2020.01021

Korir, N. K., Han, J., Shangguan, L. F., Wang, C., Kayesh, E., Zhang, Y. Y., et al. (2013). Plant variety and cultivar identification: advances and prospects. *Crit. Rev. Biotechnol.* 33, 111–125. doi: 10.3109/07388551.2012.675314

Kumar, L. S., Sawant, S. A., Gupta, V. S., and Ranjekar, P. K. (2001). Genetic variation in Indian populations of Scirpophaga incertulas as revealed by RAPD-PCR analysis. *Biochem. Genet.* 39, 43–57. doi: 10.1023/A:1002797219182

Kumari, R., Wankhede, D. P., Bajpai, A., Maurya, A., Prasad, K., Gautam, D., et al. (2019). Genome wide identification and characterization of microsatellite markers in black pepper (Piper nigrum): A valuable resource for boosting genomics applications. *PLoS One* 14:e0226002. doi: 10.1371/journal.pone.0226002

Lee, J., Izzah, N. K., Choi, B. S., Joh, H. J., Lee, S. C., Perumal, S., et al. (2015). Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (Brassica oleracea L.). *DNA Res.* 23, 29–41. doi: 10.1093/dnares/dsv034

Li, D. J., Deng, Z., Qin, B., Liu, X. H., and Men, Z. H. (2012). De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (Hevea brasiliensis Muell. Arg.). *BMC Genomics* 13:192. doi: 10.1186/1471-2164-13-192

Li, Y. C., Korol, A. B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007. doi: 10.1093/molbev/msh073

Liu, S. R., An, Y. L., Li, F. D., Li, S. J., Liu, L. L., Zhou, Q. Y., et al. (2018). Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (Camellia sinensis). *Mol. Breed.* 38:59. doi: 10.1007/s11032-018-0824-z

Liu, L., Guo, W., Zhu, X., and Zhang, T. (2003). Inheritance and fine mapping of fertility-restoration for cytoplasmic male sterility in Gossypium hirsutum L. *Theor. Appl. Genet.* 106, 461–469. doi: 10.1007/s00122-002-1084-0

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, W. C., Xu, Y. T., Li, Z. K., Fan, J., and Yang, Y. (2019). Genome-wide mining of microsatellites in king cobra (Ophiophagus hannah) and cross-species development of tetranucleotide SSR markers in Chinese cobra (Naja atra). *Mol. Biol. Rep.* 46, 6087–6098. doi: 10.1007/s11033-019-05044-7

Liu, L. W., Zhao, L. P., Gong, Y. Q., Wang, M. X., Chen, L. M., Yu, F. M., et al. (2008). DNA fingerprinting and genetic diversity analysis of late-bolting radish cultivars with RAPD, ISSR and SRAP markers. *Sci. Hort.* 116, 240–247. doi: 10.1016/j.scienta.2007.12.011

Lv, H., Fang, Z., Yang, L., Zhang, Y., Wang, Q., Liu, Y., et al. (2014). Mapping and analysis of a novel candidate Fusarium wilt resistance gene FOC1 in Brassica oleracea. *BMC Genomics* 15:1094. doi: 10.1186/1471-2164-15-1094

Lv, H., Wang, Y., Han, F., Ji, J., Fang, Z., Zhuang, M., et al. (2020). A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. *Sci. Rep.* 10:12394. doi: 10.1038/s41598-020-69389-x

Lv, H. H., Wang, Q. B., Han, F. Q., Liu, X., Fang, Z. Y., Yang, L. M., et al. (2017). Genome-wide InDel/SSR scanning reveals significant loci associated with excellent agronomic traits of a cabbage (Brassica oleracea) elite parental line'01-20′. *Sci. Rep.* 7:41696. doi: 10.1038/srep41696

Min, X. Y., Zhang, Z. S., Liu, Y. S., Wei, X. Y., Liu, Z. P., Wang, Y. R., et al. (2017). Genome-wide development of MicroRNA-based SSR markers in medicago truncatula with their transferability analysis and utilization in related legume species. *Int. J. Mol. Sci.* 18:2440. doi: 10.3390/ijms18112440

Morgante, M., Hanafey, M., and Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30, 194–200. doi: 10.1038/ng822

Nelson, D. L., Orr, H. T., and Warren, S. T. (2013). The unstable repeats-three evolving faces of neurological disease. *Neuron* 77, 825–843. doi: 10.1016/j.neuron.2013.02.022

Portis, E., Lanteri, S., Barchi, L., Portis, F., Valente, L., Toppino, L., et al. (2018). Comprehensive characterization of simple sequence repeats in eggplant (Solanum melongena L.) genome and construction of a web resource. *Front. Plant Sci.* 9:401. doi: 10.3389/fpls.2018.00401

Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S., et al. (2016). A genome-wide survey of the microsatellite content of the globe artichoke genome and the development of a web-based database. *PLoS One* 11:e0162841. doi: 10.1371/journal.pone.0162841

Rohlf, F. (2000). *NTSYS-Pc. Numerical Taxonomy and Multivariate Analysis System, Version 2.1.* New York: Exeter software.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M. A., et al. (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS One* 8:e54710. doi: 10.1371/journal.pone.0054710

Silva, P. I., Martins, A. M., Gouvea, E. G., Pessoa-Filho, M., and Ferreira, M. E. (2013). Development and validation of microsatellite markers for Brachiaria ruziziensis obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics* 14:17. doi: 10.1186/1471-2164-14-17

Sraphet, S., Boonchanawiwa, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., et al. (2011). SSR and EST-SSR based genetic linkage map of cassava (Manihot esculenta Crantz). *Theor. Appl. Genet.* 122, 1161–1170. doi: 10.1007/s00122-010-1520-5

Taheri, S., Abdullah, T. L., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., et al. (2018). Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules* 23, 1–20. doi: 10.3390/molecules23020399

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Uncu, A. T. (2019). Genome-wide identification of simple sequence repeat (SSR) markers in Capsicum chinense Jacq. With high potential for use in pepper introgression breeding. *Biologia* 74, 119–126. doi: 10.2478/s11756-018-0155-x

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trend Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005

Wang, Y. J., Li, X. Y., Han, J., Fang, W. M., Li, X. D., Wang, S. S., et al. (2011). Analysis of genetic relationships and identification of flowering-mei cultivars using EST-SSR markers developed from apricot and fruiting-mei. *Sci. Hort.* 132, 12–17. doi: 10.1016/j.scienta.2011.09.013

Wang, L. Y., Li, S. S., Wang, T. Y., He, C. Y., Luo, H. M., Zhang, J. G., et al. (2021). Genomic SSR and EST-SSR markers for phylogenetic and pedigree reconstructions-A comparison in sea buckthorn. *Plant Breed.* 140, 167–183. doi: 10.1111/pbr.12889

Xu, Y., Zeng, A., Song, L., Li, J., and Yan, J. (2019). Comparative transcriptomics analysis uncovers alternative splicing events and molecular markers in cabbage (Brassica oleracea L.). *Planta* 249, 1599–1615. doi: 10.1007/s00425-019-03108-3

Xue, H., Zhang, P., Shi, T., Yang, J., Wang, L., Wang, S., et al. (2018). Genome-wide characterization of simple sequence repeats in Pyrus bretschneideri and their application in an analysis of genetic diversity in pear. *BMC Genomics* 19:473. doi: 10.1186/s12864-018-4822-7

Yang, Y. Y., He, R. Q., Zheng, J., Hu, Z. H., Wu, J., and Leng, P. S. (2020). Development of EST-SSR markers and association mapping with floral traits in Syringa oblata. *BMC Plant Biol.* 20:436. doi: 10.1186/s12870-020-02652-5

Yang, H., Li, C., Lam, H. M., Clements, J., Yan, G., and Zhao, S. (2015). Sequencing consolidates molecular markers with plant breeding practice. *Theor. Appl. Genet.* 128, 779–795. doi: 10.1007/s00122-015-2499-8

Zhai, L. L., Xu, L., Wang, Y., Cheng, H., Chen, Y. L., Gong, Y., et al. (2014). Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (Raphanus sativus L.). *Mol. Breed.* 33, 611–624. doi: 10.1007/s11032-013-9978-x

Zhang, J. Z., Liu, S. R., and Hu, C. G. (2016). Identifying the genome-wide genetic variation between precocious trifoliate orange and its wild type and developing new markers for genetics research. *DNA Res.* 23, 403–414. doi: 10.1093/dnares/dsw017

Zhang, Z., Zhang, J. W., Yang, Q., Li, B., Zhou, W., and Wang, Z. Z. (2021). Genome survey sequencing and genetic diversity of cultivated Akebia trifoliata assessed *via* phenotypes and SSR markers. *Mol. Biol. Rep.* 48, 241–250. doi: 10.1007/s11033-020-06042-w

Zhong, H., Zhang, F., Zhou, X., Pan, M., Xu, J., Hao, J., et al. (2017). Genome-wide identification of sequence variations and SSR marker development in the Munake grape cultivar. *Front. Ecol. Evol.* 9:664835. doi: 10.3389/fevo.2021.664835

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.

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

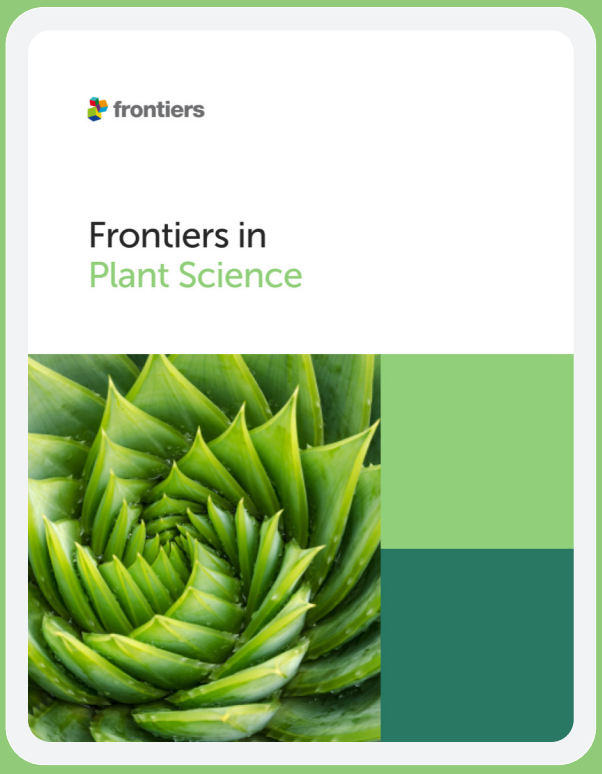
Frontiers in Plant Science

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