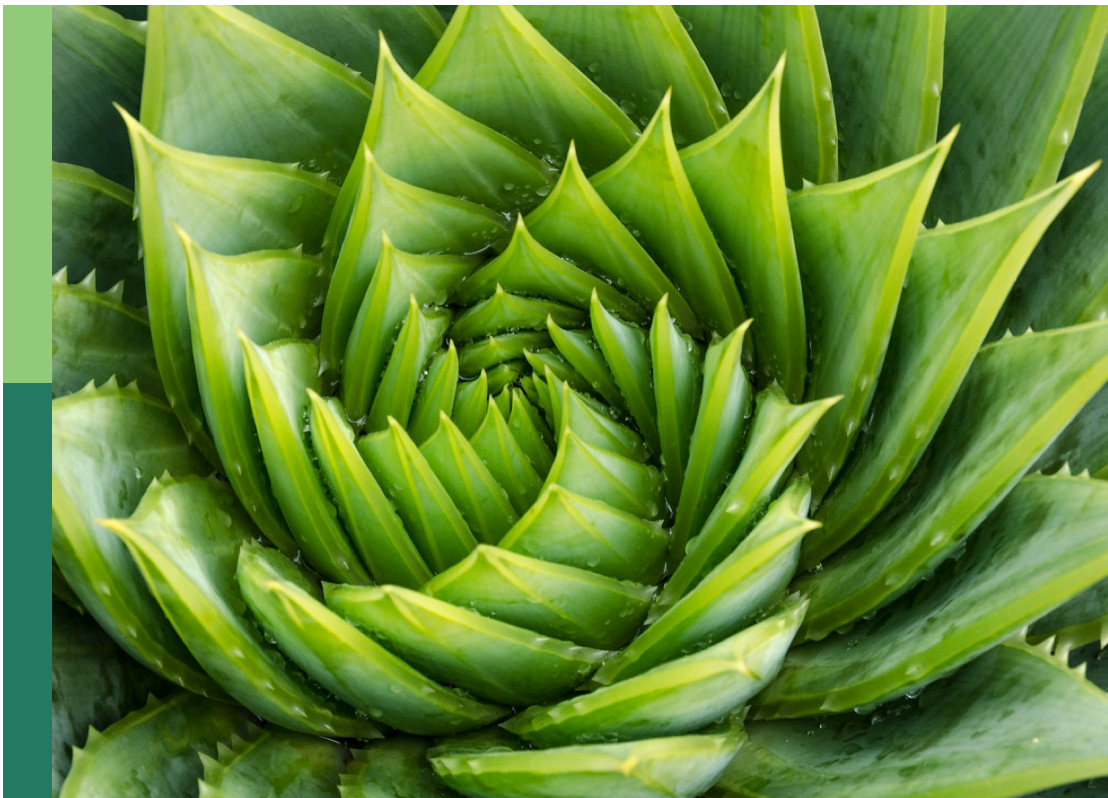
frontiers | Research Topics

Research on brassicaceae crops genomics and breeding

Edited by
Xiangshu Dong, Yoonkang Hur and Xiaodong Yang

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2985-0
DOI 10.3389/978-2-8325-2985-0

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

Research on brassicaceae crops genomics and breeding

Topic editors

Xiangshu Dong — Yunnan University, China

Yoonkang Hur — Chungnam National University, Republic of Korea

Xiaodong Yang — Yangzhou University, China

Citation

Dong, X., Hur, Y., Yang, X., eds. (2023). *Research on brassicaceae crops genomics and breeding*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2985-0

Table of contents

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fpls.2022.1019513

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Xiaochun Wei,
Henan Academy of Agricultural Sciences (HNAAS), China
Hui Feng,
Shenyang Agricultural University, China

*CORRESPONDENCE
Jungen Kang
kangjungen@nercv.org

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 15 August 2022
ACCEPTED 05 September 2022
PUBLISHED 26 September 2022

CITATION
Zhong X, Yue X, Cui J, Han R, Gao Y and Kang J (2022) Complete Mitochondrial Genome Sequencing and Identification of Candidate Genes Responsible for C5-Type Cytoplasmic Male Sterility in Cabbage (*B. oleracea* var. *capitata*).
Front. Plant Sci. 13:1019513.
doi: 10.3389/fpls.2022.1019513

Complete mitochondrial genome sequencing and identification of candidate genes responsible for C5-type cytoplasmic male sterility in cabbage (*B. oleracea* var. *capitata*)

Xionghui Zhong[1], Xiangqing Yue[1,2], Jian Cui[1], Rui Han[1,2], Yi Gao[1] and Jungen Kang[1]*

[1]Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Ministry of Agriculture, Beijing, China, [2]College of Horticulture, Gansu Agricultural University, Lanzhou, China

Cytoplasmic male sterility (CMS) is widely used in cruciferous vegetables hybrid breeding. The C5-type CMS cabbage line exhibits stable male sterility and offers great value for cabbage breeding. However, the underlying CMS mechanism remains unclear. Here, the complete mitochondrial genome was sequenced and assembled for this line. The genome size was 221,862 bp. Mitochondrial genome comparison showed that the mitochondrial genome was likely generated by recombination with a *nap*-type CMS *B. napus* strain. Sixty-seven unknown-function open reading frames (ORFs) were identified. Seven *orfs*, *orf114a*, *orf123a*, *orf188a*, *orf*222a, *orf*261a, *orf286a*, and *orf322a*, were specifically identified in this genome. The presence of these candidate CMS genes decreased ATPase activity and ATP content by affecting the transcript levels of energy metabolism-related genes and F_1F_0-ATP synthase assembly. Among them, *orf188a*, *orf222a*, *orf261a*, *orf286a*, and *orf322a* possessed a transmembrane structure, and *orf188a* was cotranscribed with *rps7* and *trnfM*. *orf222a* was partially homologous to *atp8* and coexpressed with *nad5*. *orf261a* and *orf322a* were cotranscribed with *cox1* and *atp9*, respectively. Additionally, *orf114a* was cotranscribed with *atp8*. Yeast two-hybrid assays showed that the ORF222a protein interacts with a *B. oleracea* ATP17 homolog (*Bo7g114140*) during F_0-type ATP synthase assembly, reducing the quantity and activity of assembled F_1F_0-ATP synthase. Cytological sections showed that premature separation of the tapetum from the connective tissue and delayed tapetal programmed cell death (PCD) might be the immediate causes of CMS in C5-type CMS cabbage lines. Our results may help uncover the molecular mechanism of C5-type CMS in *B. oleracea* from the perspectives of the whole mitochondrial genome and cytology of anther development.

KEYWORDS
cytoplasmic male sterility (CMS), ORF222a, ATP17, mitochondrial genome, tapetal cell

Introduction

Cabbage (*Brassica oleracea* var. *capitata*) is one of the most important vegetables and is grown on five continents around the world (Chiang et al., 1993). In the last few years, approximately 70 million tonnes of cabbage have been produced per year worldwide; according to the Food and Agriculture Organization of the United Nations (FAO), China is the largest cabbage producer in the world, producing approximately 34 million tonnes in 2020. Heterosis has been widely applied in cereal crop and vegetable production (Yu et al., 2021). Cytoplasmic male sterility (CMS) is an extranuclear maternally transmitted trait that produces either aborted or infertile pollen grains (Dong et al., 2013). CMS lines, along with maintainer and restorer lines, are highly valuable resources for F1 hybrid seed production, which has been widely developed and applied to all *Brassicaceae* crops (Yamagishi and Bhat, 2014; Yang et al., 2021).

In *Brassica* species, a variety of CMS types have been independently reported, such as *Ogu*, *nap*, *Nig*, *tour*, *Shaan2A*, *pol*, *Nsa*, *hau*, *inap*, and *oxa*. Among them, Ogu-CMS was first identified in *Raphanus sativus* (Ogura, 1967); *Nig* CMS comes from *Brassica nigra* (Pearson, 1972); *nap* CMS, *Shaan2A* CMS and *pol* CMS are derived from *Brassica napus* (Shiga and Baba, 1971; Li, 1980; Fu, 1981); *Nsa* CMS is an alloplasmic male sterility system derived from somatic hybridization between *Brassica napus* and *Sinapis arvensis* (Hu et al., 2002); *inap* CMS was obtained from somatic hybridization between *Brassica napus* and *Isatis indigotica* by recurrent backcrossing (Kang et al., 2017); and *tour* CMS, *hau* CMS and *oxa* CMS originated from *Brassica juncea* (Rawat and Anand, 1979; Wan et al., 2008; Heng et al., 2018). The *Nig*, *tour*, *pol* and *Ogu* CMSs were transferred into *B. oleracea* by protoplast fusion or the interspecific cross; however, the cybrids exhibited poor agronomic performance except for Ogu-CMS system (Pearson, 1972; Bannerot H et al., 1974; Yarrow et al., 1990; Cardi and Earle, 1997). To date, the Ogu-CMS system has been used as the major CMS system for F1 seed production in *B. oleracea* crops because of its stable sterility and lack of adverse effects on plant growth (Dey et al., 2013). However, the broad use of a single CMS results in a high degree of cytoplasmic genetic uniformity, which may lead to genetically vulnerable plants (Yamagishi and Terachi, 2017). Therefore, it is urgent to develop more CMS systems for cabbage breeding.

Thus far, many CMS-related genes have been identified and reported in *Brassica* species. Most open reading frame (ORF)-encoded proteins in CMS lines disrupt energy metabolism, which disrupts the F_1F_0-ATP synthase subunit. The molecular structure of the F_1F_0-ATP synthase complex from *Saccharomyces cerevisiae* has already been built (Srivastava and Luo, 2018). However, the structure of this complex in plants has not been resolved. Plant F_1F_0-ATP synthase shares the basic structure of the enzyme complexes described in yeast. The water soluble complex (F_1) component possesses 3 copies each of subunits α (ATP1) and β (ATP2) and one copy each of subunits γ (ATP3), δ (ATP16) and ϵ (ATP15). The membrane complex (F_0) is composed of subunits a (ATP6-1 and ATP6-2), b (ATP4), 8 (ATP8), c (ATP9), f (ATP17), g (ATP20) and e (ATP21) (Zancani et al., 2020). Subunits a (ATP6), 8 (ATP8), and c (ATP9) have been extensively investigated in most CMS types. Several previous findings showed that products of candidate CMS-related genes often affect mitochondrial functions *via* interacting with the other essential nuclear/mitochondria-encoded proteins. For examples, ORFH79 impairs mitochondrial function *via* interaction with P61, which is a subunit of electron transport chain complex III in Honglian CMS rice line (Wang et al., 2013). The mitochondrial protein WA352 interacts with COX11 to inhibit its function in peroxide metabolism, which induces the premature tapetal PCD and triggers consequent pollen abortion in the Wild Abortive CMS (CMS-WA) rice line (Luo et al., 2013). ORF224 was found to be associated with a respiratory electron transport chain protein (BnaC03g14740D), which affects the development of anthers and induces pollen abortion in the *pol* CMS line of *B. napus* (Wang et al., 2021). Additionally, the CMS-related genes can also affect mitochondrial functions though cotranscribing with the essential mitochondrial genes. The sterility gene *orf138*, which triggers Ogu-CMS, is cotranscribed with *atp8* and *trnfM* (Tanaka et al., 2012; Zhong et al., 2021). The sterility gene *orf224* is responsible for *pol* CMS, which is cotranscribed with *atp6* (Singh and Brown, 1993). *orf222*, the key gene of *nap* CMS, is similar to the *pol* CMS-associated gene *orf224* and the cotranscribed genes *nad5c* and *orf139* (Lhomme et al., 1997). *orf288*, the master gene of *hau* CMS, is also cotranscribed with the downstream gene *atp6* (Heng et al., 2014). The key sterility gene *orf346*, which controls *Nsa* CMS, is cotranscribed with the *nad3* and *rps12* complex (Sang et al., 2021). The sterility gene *orf193*, which triggers *tour* CMS (originating from cell fusion between *B. napus* and *B. tournefortii*), is cotranscribed with the downstream gene *atp9-2* (Dieterich et al., 2003). However, *orf263*, the causal gene for *tour* CMS (obtained from sexual hybridization between *B. napus* and *B. tournefortii*), is

cotranscribed with the downstream gene *atp6* (Landgren et al., 1996). In total, the expression of specific mitochondrial genes is highly dependent on the species in which the CMS source originated.

In the present study, C5-type CMS was identified as a novel alloplasmic male-sterility system derived from intergeneric hybridization of *B. napus* and *B. oleracea*. To provide new insights into the molecular mechanism of this C5-type CMS of cabbage exhibiting anther indehiscence and no pollen dispersal, we sought to identify CMS-associated candidate genes in mitochondria using a next-generation sequencing approach and accordingly analyzed the specific ORFs. The complete mitochondrial genomes of C5-type CMS and its maintainer line were sequenced. To deeply explore the abortive mechanisms of the C5-type CMS in cabbage, we investigated the characteristics of maintainer and C5-type CMS cabbage lines in cytological, physiological, and molecular analyses. To ascertain the responsible gene and the underlying mechanisms for C5-type CMS, we characterized C5-type CMS-associated candidate genes. The mitochondrial genome comparison results indicated that the C5-type CMS original cytoplasm donor material was most likely generated by recombination with the *nap*-type CMS *B. napus* strain through intergeneric hybridization during the breeding process. The CMS candidate protein ORF222a interacts with the ATP17 homolog (*Bo7g114140*) in *B. oleracea*, impairing the step of F_1Fo-ATP synthase assembly, which results in remarkable reductions in the ATP and ATPase activity levels in the anther samples of C5-type CMS cabbage lines compared with those of the maintainer lines. Cytological analysis of anther development between the maintainer and C5-type CMS cabbage lines showed that premature separation of the tapetum from the connective tissue and abnormal degradation of tapetal cells might be the immediate cause of CMS in C5-type CMS cabbage lines.

Materials and methods

Plant materials

The maintainer line was an open-pollinating and early-maturing cabbage variety. The original cytoplasmic donor cabbage material was introduced from *B. napus nap*-type cytoplasm by intergeneric hybridization. The C5-type CMS cabbage line was obtained from crosses and consecutive backcrosses with *B. napus* in 2005 by Dr. Jungen Kang from the Beijing Vegetable Research Center of the Beijing Academy of Agriculture and Forestry Sciences (BAAFS). The stability of C5-type CMS was observed for more than 10 years. The mitochondrial DNA of both the C5-type CMS and maintainer lines was sent to Biozeron Company (Shanghai, China) for DNA library construction and sequencing in 2018.

Mitochondrial genome sequencing and assembly

One microgram of purified mitochondrial DNA was fragmented for 300-500 bp paired-end library construction using a TruSeq™ Nano DNA Sample Prep Kit. Sequencing was performed on the Illumina HiSeq 4000 platform (BIOZERON Co., Ltd., Shanghai, China). A DNA library with approximately 15-20 kb SMRTbell libraries was constructed and sequenced on a PacBio Sequel Sequencer (PacBio Inc., Menlo Park, CA, USA). For the cabbage mitochondrial genome assembly, the filtered Illumina HiSeq subreads were preliminarily assembled by ABySS v2.0.2 software (version 1.5) (Jackman et al., 2017). Then, the PacBio Sequel data were aligned by the blasR method for single-molecule sequencing data correction. The corrected PacBio Sequel data and Illumina HiSeq data were combined for the mitochondrial genome framework assembly using SPAdes v3.10.1 software (Antipov et al., 2016). Finally, clean Illumina HiSeq reads were mapped to the assembled mitochondrial genome to verify the accuracy of the sequence. The circular genome maps were drawn using OrganellarGenomeDRAW (version 1.2) (Greiner et al., 2019).

Gene prediction and annotation

Mitochondrial genes were predicted based on a combination of homology-based gene prediction and *de novo* prediction by Genewise and AUGUSTUS software. The tRNA and rRNA genes were predicted using tRNAscan-SE (version 2.0) and rRNAmmer (version 1.2), respectively (Lowe and Eddy, 1997; Lagesen et al., 2007). The functions of the predicted proteins were annotated based on a BLASTP search against universal databases, such as the National Center for Biotechnology Information (NCBI) database, the Gene Ontology (GO) database (Ashburner et al., 2000), the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (eggNOG) database (Jensen et al., 2008), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004). The transmembrane domains in each candidate ORF-encoded protein were assessed using TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

ATP content and synthase activity measurement

The ATP content and synthase activity were measured in the anthers of the maintainer and C5-type CMS lines with an ATP content assay kit and a Na^+K^+-ATP synthase activity assay kit, respectively (Solarbio Co., Ltd., Beijing, China). The procedures were performed according to the manufacturer's instructions (Solarbio, BC0300 and BC0065).

Y2H assay

The full-length coding sequences (CDSs) of *atp4*, *atp6*, *atp8*, *atp9*, and two *atp17*-homologous genes (*Bo3g175820* and *Bo7g114140*) were cloned into pGADT7 vectors; *orf188a*, *orf222a* and *orf261a* were cloned into the PGBKT7 vector. The eighteen candidate interaction combinations were transformed into the yeast strain Y2HGold. Y2H assays were performed according to the Clontech manual. The transformed yeast cells were inoculated onto synthetic defined medium without leucine and tryptophan (SD/-Leu/-Trp) and incubated at 30°C for 2-3 days. The single transformation colony was then gradient-diluted and inoculated onto SD/-Leu/-Trp/-His-Ade media and cultured for 3–5 days at 30°C to observe strain growth.

Cytological analysis

Flower buds from the maintainer and C5-type CMS lines were collected at six different developmental stages and immediately fixed in formalin-aceto-alcohol (FAA) solution. Paraffin sections were then prepared according to the method of Zhong et al. (2021). Here, anther transverse sections were stained in 1% toluidine blue solution and viewed under a Leica DMR2 microscope (Leica, Wetzlar, Germany), and images were obtained with a Nikon Coolpix4200 camera (Nikon, Tokyo, Japan)

Cloning of CMS-associated candidate ORFs and subunit genes of F_1Fo-ATP synthase

Total RNA was isolated from the anthers of maintainer and C5-type CMS lines using an RNA extraction kit (RNAprep Pure Plant Kit, Tiangen, Beijing, China). First-strand cDNA was synthesized using a PrimeScript™ RT Reagent Kit (Takara, RR037A, Dalian, China). Subunits b (*atp4*), a (*atp6*), 8 (*atp8*), and c (*atp9*) were cloned from cDNA of the C5-type CMS line; however, the subunit f (ATP17) homologs (*Bo3g175820* and *Bo7g114140*) were cloned from cDNA of the maintainer line. Cloning PCR was performed using the primers listed in Supplementary Table S1. The amplified PCR products were detected using gel electrophoresis (1.5% agarose gel).

Alignment and phylogenetic analysis

Geneious Prime software was used to perform multiple sequence alignment between *B. napus* strain 51218 and our C5-type CMS mitochondrial sequences to obtain consistent sequences. Seven mitochondrial genome sequences of *B. napus* lines (*B. napus* strain SW18: AP018473 and AP018474, strain 51218: KP161618, cultivar 088018: MW348924, cultivar NY18: MW001149, and cultivar Westar: AP006444) were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/genome/browse/). Mitochondrial genome-wide alignments and ML bootstrap analysis with 500 replicates were performed using MEGA 6.0 (Tamura et al., 2013).

Results

Flower morphology of the C5-type CMS line and maintainer lines

The sterile flowers of the C5-type CMS line were significantly smaller than the fertile flowers of the maintainer line. The filaments and anthers of fertile flowers were remarkably longer than those of sterile flowers. There were no differences in the pistils between the C5-type CMS and maintainer lines. The C5-type CMS line exhibited a sterile phenotype due to anther indehiscence and no pollen dispersal. However, the maintainer line produced normal pollen grains (Figure 1).

Mitochondrial genome sequencing and annotation of the C5-type CMS cabbage line

The mitochondrial genomes of the C5-type CMS cabbage line were sequenced with Illumina HiSeq and PacBio Sequel techniques. The Illumina HiSeq sequencing platform generated 6791 Mb of clean data. Then, 47.85 Mb of subread bases were produced for the sample in the PacBio sequencing platform. The mitochondrial genome of the C5-type CMS cabbage line was assembled into a single circular mapping molecule with a size of 221,862 bp (GenBank no. ON960289) (Figure 2). The G+C content of the C5-type CMS mitochondrial genome was 45.19%, which was comparable to that of our previously released mitochondrial genomes from maintainer and Ogura-type CMS cabbage lines. Additionally, the C5-type CMS mitochondrial genome possessed a total of 125 genes, including 99 protein-coding genes (32 known genes and 67 unknown-function ORFs), 3 rRNA genes (5S, 18S, and 26S), and 23 tRNA genes. A comparison between the maintainer and C5-type CMS mitochondrial genomes revealed that most known mitochondrial genes were identical. The C5-type CMS mitochondrial genome comprised three genes encoding the large subunits of ribosomal proteins (*rpl2*, *rpl5*, and *rpl16*), four genes encoding the small subunits of ribosomal proteins (*rps4*, *rps7*, *rps12*, and *rps14*), nine subunits of NADH dehydrogenase (*nad1*, *nad 2*, *nad* 3, *nad4*, *nad4L*, *nad*5, *nad*6, *nad7*, and *nad*9), three cytochrome oxidase subunits (*cox1*, *cox2*, and *cox3*), one cytochrome b (*cob*), five subunits of ATP synthase (*atp1*, *atp4*, *atp6*, *atp8*, and *atp9*), five cytochrome C

10.3389/fpls.2022.1019513

FIGURE 1
Flower morphology of maintainer lines and C5-type CMS lines. **(A)** Intact flowers of the C5-type CMS line and maintainer lines. **(B)** Partially dissected maintainer flower. **(C)** Partially dissected C5-type CMS flower.

synthesis-related genes (*ccmB*, *ccmC*, *ccmFN1*, *ccmFN2*, and *ccmFC*), and one maturase gene (*matR*).

Collinearity analysis of the mitochondrial genomes

To further analyze the structure and composition of the C5-type CMS cabbage mitochondrial genome, collinearity analysis was performed with the maintainer line. Our sequencing results showed that the mitochondrial genome size of the C5-type CMS line (221,862 bp) was 1,900 bp larger than that of its maintainer line (219,962 bp). In addition, a total of 9 syntenic regions (named blocks 1-9) were identified in the mitochondrial genomes of the maintainer and C5-type CMS cabbage lines on the basis of sequence homology (Figure 3A). These regions ranged from 2,317 to 66,388 bp in length and accounted for 95.85% of the mitochondrial genome sequence (Supplementary Table S2). The corresponding blocks between C5-CMS and its maintainer line had at least 99.22% identity (Supplementary Table S2). Although all 9 syntenic regions shared high sequence similarity, the directions and positions of these syntenic regions differed between the two mitochondrial genomes. The orientation of seven regions was identical, but in the other two (block 2 and block 3), the blocks were inversions. Block 4, block 8, and block 9 were translocations (Figure 3). The syntenic regions were largely discrepant in the two genomic positions, although the genetic sequences were well conserved. Recombination and rearrangement events during the breeding

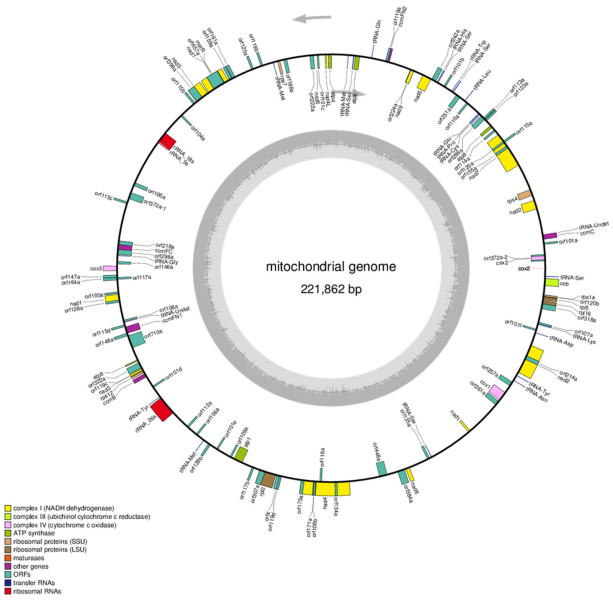


FIGURE 2
Mitochondrial genome map of the C5-type CMS line. Genes with names inside the circle are transcribed clockwise. Genes with names outside the circle are transcribed counterclockwise. The colors of the genes denote the functions of the gene products.

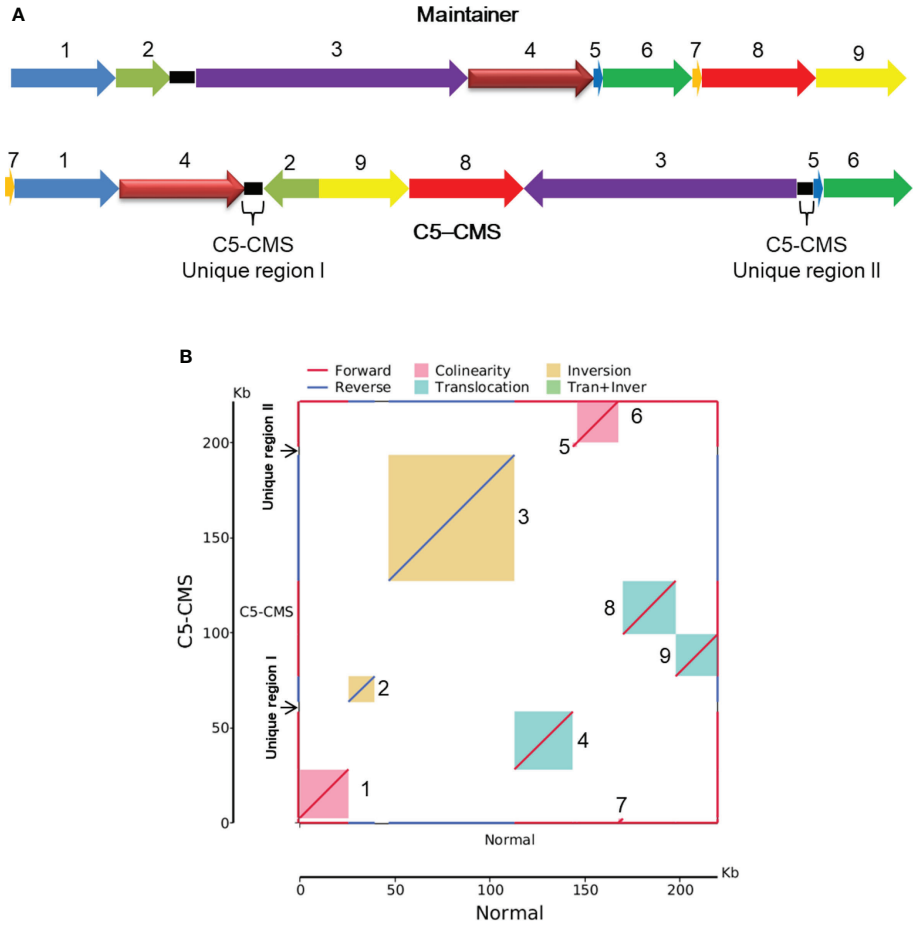


FIGURE 3
Collinearity analysis of the mitochondrial genomes between the maintainer and C5-type CMS lines. **(A)** Schematic illustration of 9 syntenic regions in the mitochondrial genomes of the maintainer and C5-CMS cabbage lines. The nine syntenic regions were named block 1 to block 9. The C5-CMS line had two unique regions. **(B)**. Maintainer line genome (X-axis) plotted against the C5-CMS line genome (Y-axis). Red lines indicate forward alignment, and blue lines indicate reverse alignment. The numbers behind the inner squares for the syntenic regions correspond to those indicated in panel **(A)** The color of the inner bars represents the alignment types. Red color: collinearity; green color: translocation; yellow color: inversion.

process were determined to be responsible for the structural differences in the mitochondrial genomes of the same species. Block 4/block 2 and block 3/block 5 in the C5-CMS mitochondrial genome were broken by unique region I and unique region II, which were nonhomologous to the regions in the maintainer mitochondrial genome (Figure 3).

Analysis of the unique regions in the C5-type CMS mitochondrial genome and phylogenetic analysis of mitochondrial genomes

The collinearity analysis results revealed two unique regions in the C5-type CMS mitochondrial genome (Figure 3A). The sizes of unique region I and unique region II were 4,897 bp and 4,141 bp, respectively. The total length of the unique regions was 9,038 bp, accounting for 4.07% of the whole C5-type CMS mitochondrial genome. At the nucleotide level, BLASTn analysis showed that the unique regions shared high similarity with mitochondrial sequences of *B. napus* strain SW18 (AP018473 and AP018474), strain 51218 (KP161618), cultivar 088018 (MW348924), cultivar NY18 (MW001149), cultivar 56366 (KM454975), and cultivar Westar (AP006444) (Figures 4A, B). The results suggested that the C5-type CMS cytoplasm may have been derived from certain *B. napus* CMS haplotypes. Furthermore, phylogenetic analysis was performed using eight complete mitochondrial genomes. Whole-genome-wide alignments and maximum likelihood (ML) bootstrap analysis with 500 replicates were

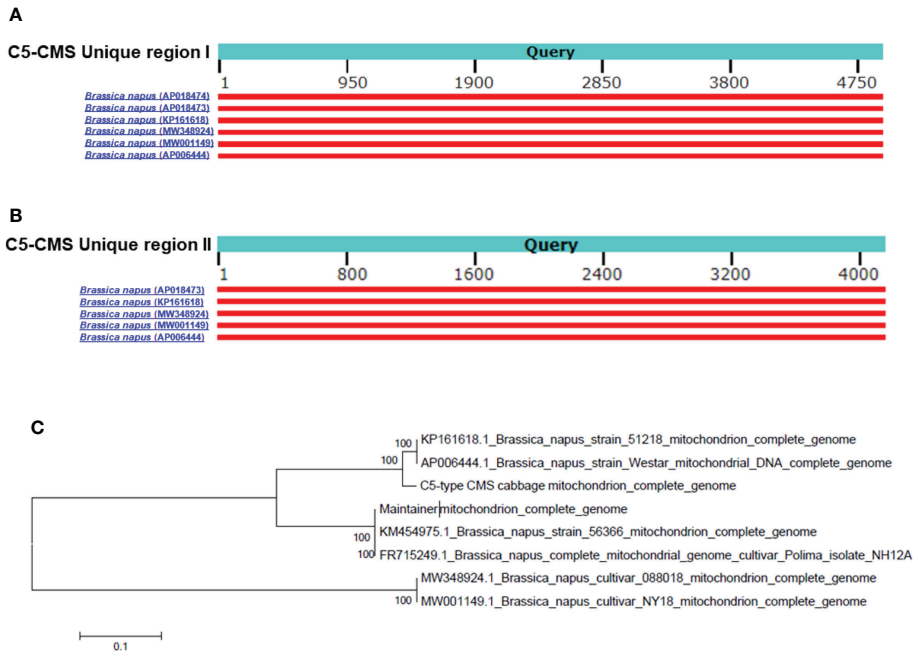


FIGURE 4
C5-CMS unique regions homologous to mitochondrial genomes and phylogenetic analysis based on the mitochondrial genome. Alignment of the C5-CMS unique region I **(A)** and C5-CMS unique region II **(B)** to the mitochondrial genomes of the *B. napus* lines (AP018474, AP018473, KP161618, MW348924, MW001149, and AP006444). The query sequences were C5-CMS unique region I (4897 bp) and unique region II (4141 bp). Red-colored boxes indicate the scores of sequence alignment with the mitochondrial genomes. **(C)** Phylogenetic tree constructed using C5-type CMS mitochondrial sequence and seven complete mitochondrial genome sequences of *B. napus* lines (AP018474, AP018473, KP161618, MW348924, MW001149, KM454975 and AP006444). The unrooted phylogenetic tree was created in MEGA 6.0 software by the ML method with 500 bootstrap iterations.

performedusing MEGA 6.0. As shown in Figure 4C, the phylogenetic tree revealed the topological structure of eight selected taxa, which were mainly divided into three clades. The C5-type CMS mitochondrial sequence formed Clade I with *B. napus* strains 51218 and Westar, which are *nap*-type strains of *B. napus*. Clade II comprised the maintainer cabbage line and *B. napus* strains 56366 and NH12A, which are *pol*-type *B. napus*. Another two *nap*-type *B. napus* cultivars, 088018 and NY18, constituted Clade III. Four *nap*-type accessions were clustered into two different clades, indicating that the investigated *nap*-type *B. napus* strains have a polyphyletic maternal origin. To further compare the mitochondrial genomes between C5-type CMS and *B. napus* cultivars 51218 and Westar, dot plot analysis and pairwise alignment were performed with Geneious Prime software. The results showed that the sequence of the C5-type CMS mitochondrial genome was more homologous to that of the *B. napus* cultivar 51218 than to that of Westar. However, there was an inversion located at the position from 80,500 bp to 93,000 bp, and many single-nucleotide polymorphisms (SNPs) were distributed throughout the genome (Figure 5). Taken together, these results show that the C5-type CMS original cytoplasm donor material was most likely generated by intergeneric hybridizations with certain *nap*-type CMS *B. napus* strains during the breeding process.

Identification of CMS-associated ORFs in the C5-type CMS mitochondrial genome

To reveal the genes determining the CMS phenomenon, we compared the mitochondrial genomes between the C5-type CMS line and its maintainer lines. Genome rearrangement events may have caused the presence of variation in genome sequence segments. Specific genes were generated by recombination, which is generally responsible for the CMS trait. Seven ORFs that encoded over 100 amino acids, including *orf114a*, *orf123a*, *orf188a*, *orf222a*, *orf261a*, *orf286a*, and *orf322a*, were specifically identified in the C5-type CMS mitochondrial genome (Table 1). All of these specific ORFs were candidate CMS genes. To further verify whether the proteins encoded by these ORFs were candidate proteins for CMS, the structures of these three proteins were predicted. We found that ORF188a, ORF222a, ORF261a, ORF286a, and ORF322a had transmembrane domains, whereas ORF114a and ORF 123a lacked transmembrane domains (Figure 6).

10.3389/fpls.2022.1019513

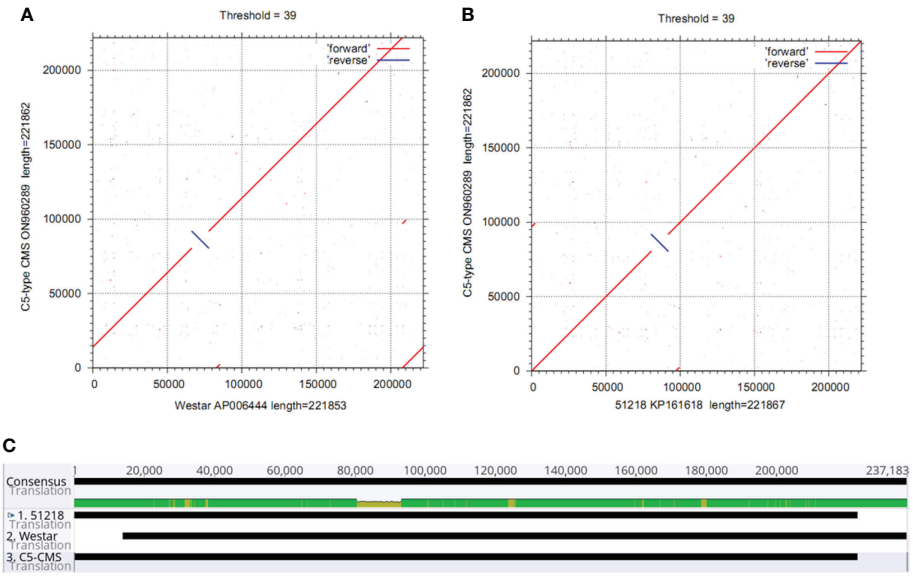


FIGURE 5
Comparative analysis of the mitochondrial genomes between the C5-type CMS line and the *B. napus* cultivars 51218 and Westar. **(A)** Dot plot alignment of the cabbage C5-type CMS and *B. napus* cultivar 51218 mitochondrial genomes. The comparisons show that the mitochondrial genome nucleotide sequences of *B. napus* cultivar 51218 (horizontal axis) resemble those of C5-type CMS cabbage (vertical axis). **(B)** Dot plot alignment of the C5-type CMS (vertical axis) and Westar (horizontal axis) mitochondrial genomes. The numbers indicate syntenic regions using the 51218 and Westar genome sequences as references. Comparisons showed that the mitochondrial genome sequence of the C5-type CMS line was similar to that of the *B. napus* cultivar 51218. **(C)** Pairwise alignment of mitochondrial genomes between the C5-type CMS line and *B. napus* cultivars 51218 and Westar was performed using Geneious Prime software. The sequence of the C5-type CMS mitochondrial genome was more homologous to that of 51218 than to that of Westar.

Organization of mitochondrial genome regions associated with candidate sterility genes for C5-type CMS

A more detailed analysis of the organization of mitochondrial genome regions associated with candidate sterility genes for C5-CMS was carried out. *orf188a* was located in C5-type CMS unique region I 651 bp downstream of *rps7*, and *orf222a* was located between C5-CMS block 4 and unique region I (Supplementary Table S3 and Figure 7A); however, *orf261a* was located 219 bp downstream of *cox1* between C5-CMS unique region II and block 5 (Supplementary Table S3 and Figure 7A). *orf322a* was located 311 bp downstream of *atp9* between block 8 and block 3, and *orf114a* and *orf123a* were located 465 bp downstream and 961 bp upstream of *atp8* within block 1, respectively (Supplementary Table S3, Figures 7B, C). However, *orf286a* was located 1,324 bp upstream of *atp8* between block 1 and block 4 of the C5-type CMS mitochondrial genome. Taken together, the findings indicate that the CMS candidate genes *orf114a*, *orf188a*,

TABLE 1 Specific ORFs in the Ogura-CMS mitochondrial genome.

Specific ORFs	Most similar mitochondrial sequence of another species	Location in the C5-type CMS line
orf114a	YP_717103.1 hypothetical protein BrnapMp004 [*Brassica napus*]	Block 1
orf123a	AKD00165.1 hypothetical protein [*Brassica napus*]	Block 1
orf188a	YP_717121.1 hypothetical protein BrnapMp023 [*Brassica napus*]	Unique region I
orf222a	YP_717120.1 hypothetical protein BrnapMp022 [*Brassica napus*]	Between block 4 and unique region I
orf261a	YP_717164.1 hypothetical protein BrnapMp067 [*Brassica napus*]	Between unique region II and block 5
orf286a	YP_717106.1 hypothetical protein BrnapMp007 [Brassica napus]	Between block 1 and block 4
orf322a	YP_717145.1 hypothetical protein BrnapMp048 [*Brassica napus*]	Between block 8 and block 3

Seven ORFs, including ORF114a, ORF123a, ORF188a, ORF222a, ORF261a, ORF286a, and ORF322a, were specifically identified in the C5-type CMS mitochondrial genome. They are identical to the mitochondrial proteins of Brassica napus. ORF144a and ORF123a were located in C5-type CMS block 1, ORF188a was located in C5-type CMS unique region I, ORF222a was located between C5-type CMS block 4 and unique region I, ORF261a was located between C5-type CMS unique region II and block 5, ORF286a was located between block 1 and block 4, and ORF322a was located between block 8 and block 3.

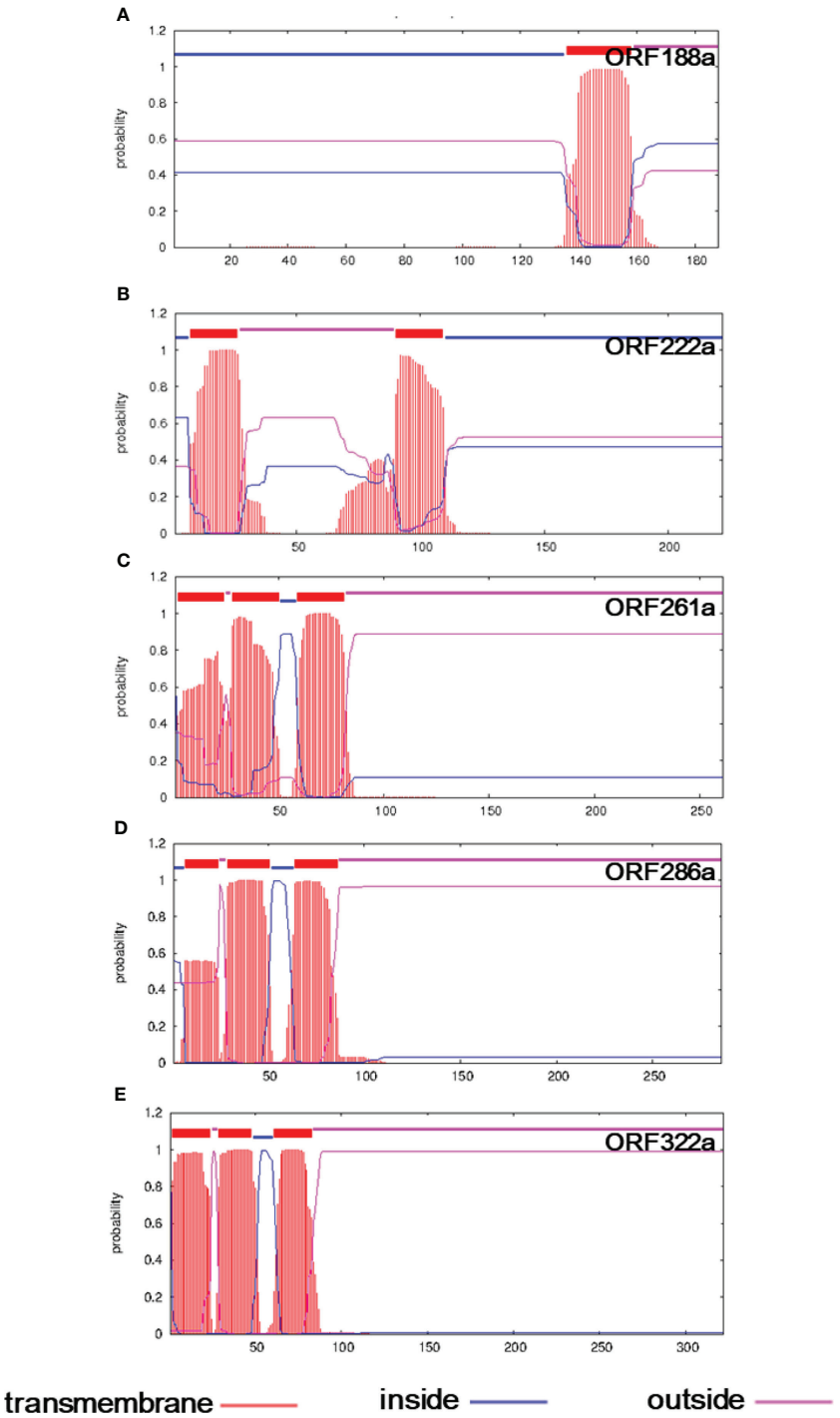


FIGURE 6
Transmembrane domain prediction of candidate CMS gene-encoded proteins in the C5-CMS line. The output of the TMHMM server indicated the location and probability associated with the predicted transmembrane domains in the C5-CMS line. **(A)** ORF188a, **(B)** ORF222a, **(C)** ORF261a, **(D)** ORF286a, and **(E)** ORF322a.

10.3389/fpls.2022.1019513

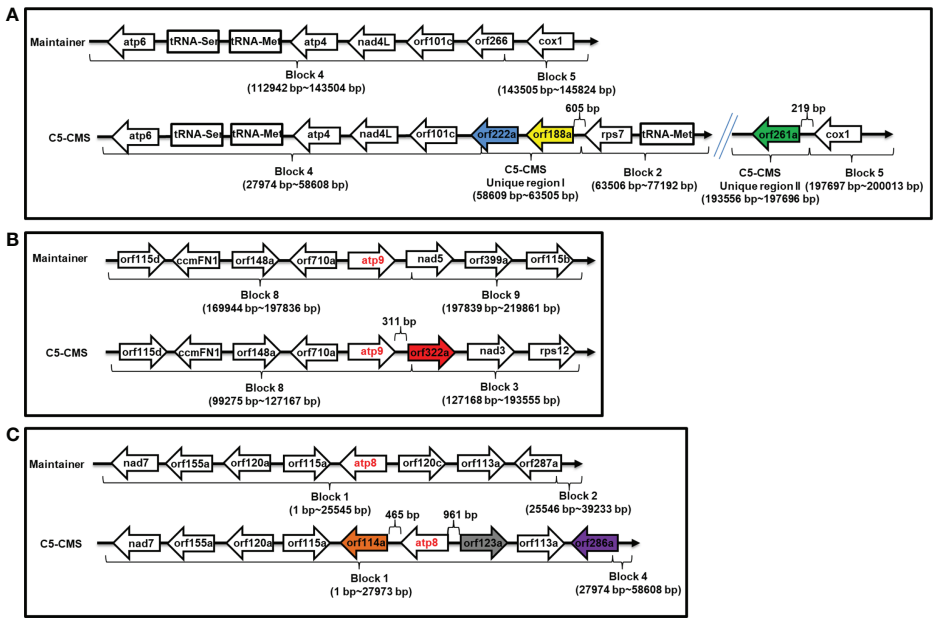


FIGURE 7
Organization of mitochondrial genome regions associated with candidate sterility genes for C5-type CMS. **(A)** Genome structure of the orf188a, orf222a, and orf261a regions. **(B)** Genome structure of the atp9 and orf322a regions. **(C)** Genome structure of the atp8, orf114a, orf123a and orf286a regions. The corresponding genes within the regions between the maintainer and C5-type CMS genomes areindicated by boxes. Two recombination events occurred in the C5-type CMS genome. Unique region I was inserted between the block 4 and block 2 syntenic regions, and orf188a and orf222a are located in unique region I. Unique region II was the result of recombination between the block 3 and block 5 syntenic regions, and orf261a is located in this region.

orf261a, and *orf322a* are cotranscribed with *atp8*, *rps7*, *cox1*, and *atp9*, respectively.

ORF222a affects F_1F_0-ATP synthase assembly by interacting with an ATP17 homolog (ATP17-2, *Bo7g114140*) in *B. oleracea*

To determine whether candidate CMS-associated ORF-encoded proteins (ORF188a, ORF222a and ORF261a) were capable of binding to subunits of F_1F_0-ATP synthase, a yeast two-hybrid (Y2H) assay was performed with ORF-encoded proteins (ORF188a, ORF222a, and ORF261a) and subunits (ATP4, ATP6, ATP8, ATP9, and ATP17) of F-ATP synthase. *orf188a*, *orf222a*, and *orf261a* were cloned into the PGBKT7 vector; meanwhile, *atp4*, *atp6*, *atp8*, *atp9*, and two *atp17*-homologous genes (*atp17-1*, *Bo3g175820* and *atp17-2*, *Bo7g114140*) were cloned into PGADT7 vectors. Eighteen candidate interaction combinations were tested with the Y2H assays. The results showed that only ORF222a interacted with the ATP17 homolog (ATP17-2, *Bo7g114140*) in *B. oleracea* (Figure 8 and Supplementary Figure S2). We thus conclude that ORF222a impairs the step of F_1Fo-ATP synthase assembly by interacting strongly with the ATP17 homolog (ATP17-2, *Bo7g114140*) in *B. oleracea*.

Detection of ATP production in the C5-type CMS cabbage line

From our whole-mitochondrial genome data, we found that the C5-type CMS-related gene *orf188a* encodes the *atp6*-like gene and that *orf222a* encodes 58 amino acids identical to ATP8 at the N-terminus (Supplementary Figure S1). We also found that *orf188a*, *orf261a*, and *orf322a* are cotranscribed with *rps7*, *cox1*, and *atp9*, respectively. Furthermore, ORF222a interacts with an ATP17 homolog (*Bo7g114140*) in *B. oleracea*. Therefore, we wondered whether these events influence the yield of ATP. We measured the ATP content and ATPase activity in anther samples from maintainer and C5-type CMS cabbage lines. We found that the ATP content and ATPase activity in the anther samples of the C5-type CMS cabbage line was remarkably lower than that in the maintainer line (Figure 9). The change in ATP yield may have been caused by altered ATPase activity. In conclusion, CMS-related genes encode partial fragments identical to mitochondrial genes with known functions and are cotranscribed with or interact with these mitochondrial electron

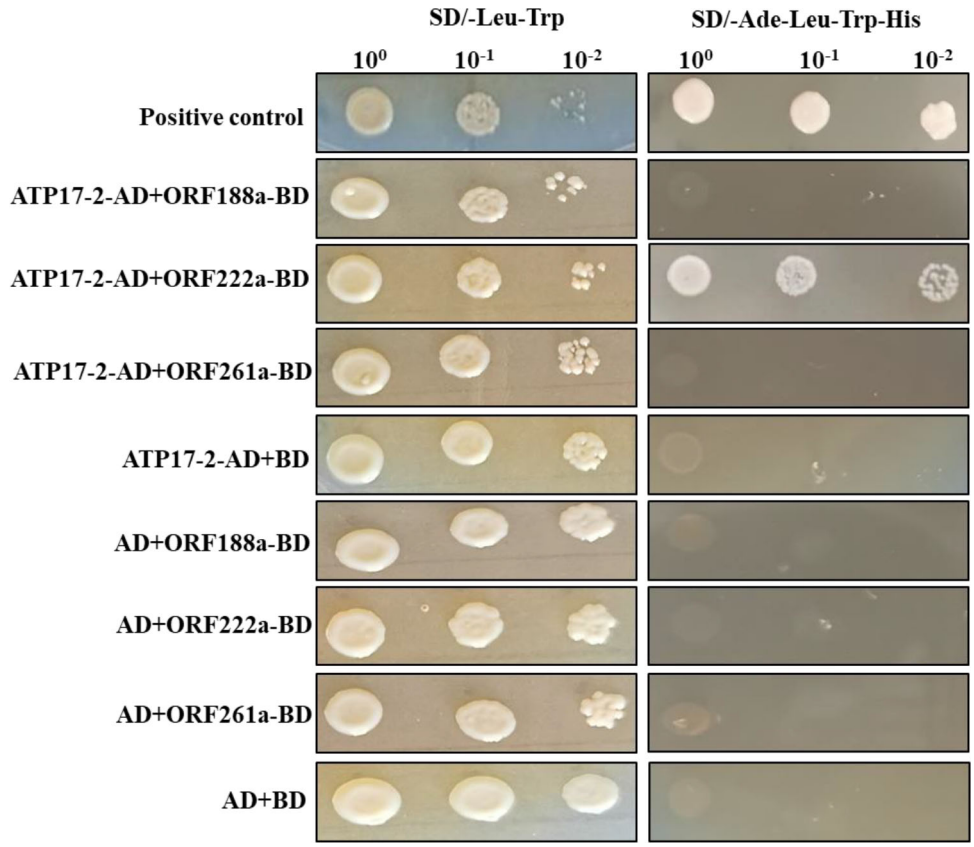


FIGURE 8
Y2H assay of the interaction of ORF222a with ATP17-2. Serial dilutions of cotransformed yeast were spotted on synthetic dropout SD/-Trp-Leu medium to select for cotransformants and SD/-Trp-Leu-His-Ade medium to select for positive interactions. p53 was used as a positive control.

transport chain (mtETC) genes, which might disturb mitochondrial energy metabolic pathways.

Cytological analysis of anther development between the maintainer and C5-type CMS cabbage lines

Our mitochondrial sequence data indicated that the C5-type CMS cabbage line possesses CMS genes from the *nap*-type CMS system of *B. napus*. However, the mechanism by which male sterility genes induce pollen abortion in cabbage is still uncertain. Thus, comparative cytological analysis of anther development between the maintainer and C5-type CMS cabbage lines was performed to address which developmental stage was affected in the stamens of the C5-type CMS line and to further determine the cause of C5-type CMS in cabbage. In the maintainer line, primordial anther tissue with four microsporangiums was found to differentiate into sporogenous cells (Figure 10A), which differentiated into irregularly shaped microspore mother cells. The tapetal cells, middle layer, endothecium and epidermis differentiated from the surrounding connective tissue (Figure 10B). Tetrads of microspores were generated after meiotic divisions, the tapetal cells began to degrade (marked by the red arrow) (Figure 10C). Free haploid microspores dissociated from the tetrads during the uninucleate microspore stage, and the most of tapetal cells degrade (Figure 10D). The microspores differentiated into mature pollen grains, and the tapetal cells degenerated completely (Figure 10E). Finally, the mature pollen grains were released from the dehiscent anther (Figure 10F). In the maintainer cabbage line, four locules of anthers developed symmetrically (Figures 10A–F). In contrast, a loss of synchronous locule development was observed in the C5-type CMS anthers during all anther development stages (marked by the black arrows) (Figures 10G–L). There were no morphological differences in the developed locules between the maintainer and C5-type CMS cabbage lines at the sporogenesis cell stage and microspore mother cell stage (Figures 10A, B, 10G, H). However, the tapetum prematurely separated from the locule wall at the tetrad stage in the C5-type CMS line (marked by the blue arrows) (Figure 10I), rather than separating from the

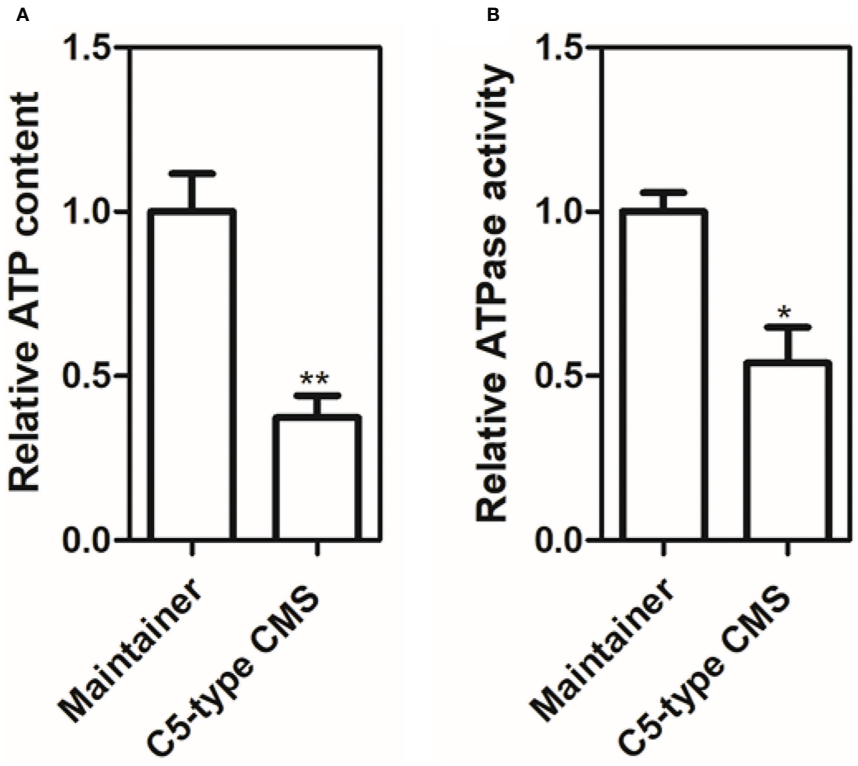


FIGURE 9
Production of ATP and ATP synthase activity in anther samples of maintainer and Ogura-CMS cabbage lines. **(A)** Comparison of the ATP content in petals of the maintainer and C5-type CMS cabbage lines. **(B)** Comparison of ATP synthase activity in petals of the maintainer and C5-type CMS cabbage lines. The *t* test comparison statistics are shown (means ± SDs; n = 3). *P < 0.05; **P < 0.01.

connective tissue during the uninucleate microspore stage, as in the maintainer cabbage line (Figure 10D). In the uninucleate microspore stage of the C5-type CMS line, haploid microspores formed but crowded together (Figure 10J). The tapetal cells were still quite intact at the uninucleate microspore stage and mature pollen stage in the C5-type CMS line (marked by the blue arrows) (Figures 10J, K). Premature separation of the tapetum from the connective tissue and delayed cell collapse of tapetal cells hindered the release of haploid microspores, which resulted in the structural collapse of microspores. The residues of microspores that failed to develop were visible, and the epidermis, fibrous layer, and intermediate layer of pollen sacs were tightly connected, generally without dehiscence (Figures 10K, L). The tapetal cellular debris was still observed at the anthesis stage (marked by the blue arrow) (Figure 10L). Taken together, these findings indicate that premature separation of the tapetum from the locule wall and abnormal degradation of tapetal cells lead to overlap of microspores, which might be the cause of CMS in C5-type CMS cabbage lines.

Discussion

Mitochondrial origin of *nap*-type CMS in C5-type CMS *B. oleracea*

Most CMS systems of *B. oleracea* (cabbage) were transferred from other cruciferous crops. Because of incomplete pollen abortion and abnormal flower organ growth, the *pol* and *Nig* CMS systems cannot be used for cabbage breeding. The Ogu-CMS system is the main type of male sterility system in cabbage breeding (Ji et al., 2020). However, a high degree of cytoplasmic genetic uniformity may lead to genetically vulnerable plants (Yamagishi and Terachi, 2017), which prompted us to generate and apply a new type of CMS in cabbage. In this study, we generated the C5-type CMS cabbage line *via* intergeneric hybridization with *nap*-type CMS *B. napus* and then consecutive backcrossing. Sequencing revealed that the mitochondrial genome of the C5-type CMS cabbage line is quite homologous to that of the *nap*-type CMS *B. napus* strains 51218 and Westar but significantly different from that

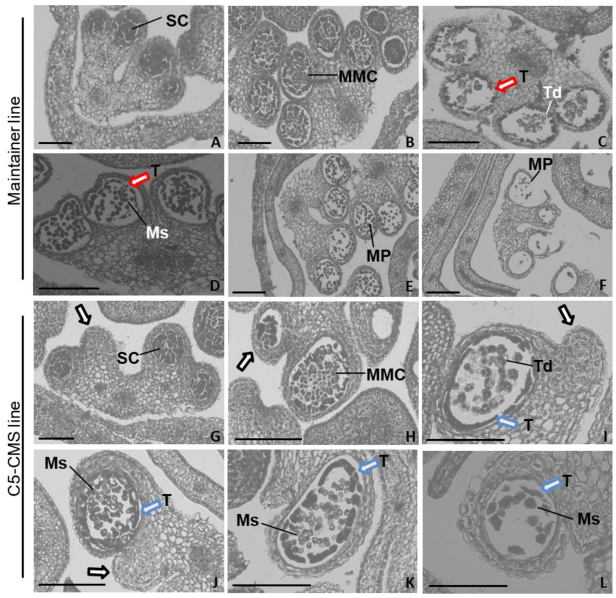


FIGURE 10
Comparative histology of anther development in *B. oleracea* maintainer and C5-type CMS lines. Histology of anther development in the maintainer line **(A–F)** and C5-type CMS line **(G–L)**. **(A, G)** are the sporogenesis cell stage; **(B, H)** are the microspore mother cell stage; **(C, I)** are the tetrad stage; **(D, J)** are the uninucleate microspore stage; **(E, K)** are the mature pollen stage; and **(F, L)** are the anthesis stage. SC, sporogenesis cell; MMC, microspore mother cells; T, tapetum; Td, tetrad; Ms, microspore; MP, mature pollen. Bar = 200 um.

of common cabbage (Figures 3B, 4C, 5). Both the chloroplast and mitochondrial genomes of progeny by sexual propagation are usually identical to those of maternal parents without recombination (Park et al., 2021). Here, the original cytoplasm of C5-type CMS cabbage was inherited from a certain *nap*-like CMS *B. napus* haplotype. However, the mitochondrial genome sequence of this *B. napus* haplotype has not yet been released.

C5-type CMS is associated with energy deficiency

Many ORFs in CMS lines have been found to disrupt energy metabolism through cotranscription with the F_1F_0-ATP synthase subunit. Here, we found that *orf188a*, *orf322a*, and *orf322a* were cotranscribed with *rps7*, *cox1*, and *atp9*, respectively (Figure 7). Most CMS-related genes have been found to be chimeric ORFs consisting of fragments of coding sequences of known mitochondrial genes from mtETC pathways, including *cox1*, *atp6*, *atp8*, *atp9*, and some other mitochondrial genes (Chen and Liu, 2014). For instance, *orf463a* for radish DCGMS-type CMS possesses a 128 bp fragment corresponding to part of the *cox1* sequence (Wang et al., 2020). *orf346* for *Nsa*-type CMS also shows partial sequence identity with the *cox1* gene (Sang et al., 2021). In our previous study, we found that the Ogura-type CMS *B. oleracea* gene *orf138*, which originated from radish, encodes an *atp8*-like gene. The CMS-related gene *orf154a* is partially homologous to the *ATP synthase subunit 1* (*atpA*) gene (Zhong et al., 2021). In *nap*-type CMS *Brassica*, chimeric *orf222* encodes a protein with a segment of ATP8 at the N-terminus plus a sequence of unknown origin (Geddy et al., 2005). *orf288* for *Hau*-type CMS *B. juncea* contains a 94 bp sequence with partial sequence homology to *nad5* and partial sequence identity to *atp9* (Heng et al., 2014). In our results, we found that the C5-type CMS-related gene *orf188a* encodes the *atp6*-like gene and that *orf222a* encodes 58 amino acids identical to ATP8 at the N-terminus (Supplementary Figure S1). The CMS-related genes encode partial fragments consistent with known functional mitochondrial genes. It has been speculated that these chimeric genes and the cotranscription events may disrupt mitochondrial functions by interfering with the expression of native mitochondrial genes, affecting the assembly or activity of different mitochondrial complexes (Shaya et al., 2012). In addition, it has been reported that chimeric genes confer the CMS phenotype by directly interacting with genes involved in the mtETC or ATP synthase complex. In maize, the chimeric gene *atp6c* confers CMS by interacting with ATP8 and ATP9, which impairs the assembly of the mitochondrial ATP synthase complex. Ultimately, this results in reductions in the quantity and activity of assembled ATP synthase required for anther development (Yang et al., 2022). In our study, we found that ORF222a impaired the step of F_1F_0-ATP synthase assembly by interacting strongly with an ATP17 homolog (*Bo7g114140*) in *B.*

oleracea (Figure 8 and Supplementary Figure S2). We also found that the ATP and ATPase activity levels in the anther samples of the C5-type CMS cabbage line were remarkably lower than those in the maintainer line (Figure 9). In addition, cytological sections showed that four locules develop synchronously through all stages of maintainer anther development (Figures 10A–F). In contrast, C5-type CMS anthers can display a loss of synchronous locule development (Figures 10G–J), which could be triggered by energy deficiency. Taken together, all of these findings indicate that C5-type CMS is associated with energy deficiency.

The abnormal development of tapetal cells probably causes male sterility in the C5-type CMS cabbage line

Programmed cell death (PCD) of the tapetum is one of the most critical steps for fertility. It is generally assumed that abnormal tapetal cells fail to produce enough nutrients and enzymes for microspore development and release from the tetrad, which often triggers the abortion of microspores (Liu et al., 2015). Additionally, defective tapetum PCD has previously been reported to cause anther indehiscence in an autophagy-deficient mutant of rice (Kurusu et al., 2014). In our results, the premature separation of the tapetum from the connective tissue and delayed cell collapse of tapetal cells hindered the release of haploid microspores, which directly resulted in both the structural collapse of microspores and anther indehiscence in the C5-type CMS line (Figure 10). This result is consistent with the development of *nap*-type CMS anthers in the *Brassica napus* (Geddy et al., 2005). However, it is quite different from the case in the Ogura-type CMS cabbage line, in which abnormal proliferation of tapetal cells hinders and destroys the development of haploid microspores through spatial constriction (Zhong et al., 2021).

Conclusion

We generated a new type of CMS cabbage line, named the C5-type CMS line, which showed abortion of microspores and anther indehiscence. *orf188a*, *orf222a*, *orf261a*, *orf286a*, and *orf 322a* were specifically identified as candidate CMS genes *via* sequencing of the mitochondrial genome of the C5-type CMS cabbage line. We found that *orf114a*, *orf188a*, *orf261a* and *orf322a* were cotranscribed with *atp8*, *rps7*, *cox1*, and *atp9*, respectively (Figure 7). The most interesting finding was that ORF222a interacts with an ATP17 homolog (*Bo7g114140*) in *B. oleracea*, which might impair the step of F_1Fo-ATP synthase assembly. We also found that the ATP and ATPase activity levels in the anther samples of the C5-type CMS cabbage line were remarkably lower than those in the maintainer line. These candidate CMS genes may perturb ATP synthesis during the anther development, which triggers a loss of synchronous locule development, premature separation of the tapetum from the connective tissue, and delayed cell collapse of tapetal cell in anthers, finally leading to CMS.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number ON960289. The data will be released on Sep 17, 2022. We also provide confirmation of deposition.

Author contributions

XZ carried out the sequence data analysis, and drafted the manuscript. XZ, XY, JC, RH and YG designed and coordinated all experiments. JK supervised the work and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (31972408) and the Science and Technology Innovation Capacity Building Projects of the Beijing Academy of Agriculture and Forestry Sciences (KJCX20200410, KJCX20210425, and KJCX20200113).

Acknowledgments

We thank Dr. Qingbiao Wang from Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, for his helpful suggestion for our 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.2022.1019513/full#supplementary-material

References

Antipov, D., Korobeynikov, A., McLean, J. S., and Pevzner, P. A. (2016). hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics.* 32 (7), 1009–1015. doi: 10.1093/bioinformatics/btv688

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 (1), 25–29. doi: 10.1038/75556

Bannerot H, B. L., Couderon, Y., and Temple, J. (1974). "Transfer of cytoplasmic male sterility from raphanus sativus to brassica oleracea," in *Proc eucarpia meet cruciferae*. Eds. A. B. Wills and C. North 52–54. Scottish Hortic Res Inst, Invergavrie, UK

Cardi, T., and Earle, E. D. (1997). Production of new CMS brassica oleracea by transfer of `Anand' cytoplasm from b. rapa through protoplast fusion. *Theor. Appl. Genet.* 94 (2), 204–212. doi: 10.1007/s001220050401

Chen, L., and Liu, Y.-G. (2014). Male Sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* 65 (1), 579–606. doi: 10.1146/annurev-arplant-050213-040119

Chiang, M. S., Chong, C., Landry, B. S., and CrÊTe, R. (1993). "8 - cabbage: Brassica oleracea subsp. capitata l," in *Genetic improvement of vegetable crops. Amsterdam: Pergamon*. Eds. G. Kalloo and B. O. Bergh 113–155. Chiang, Pergamon, UK.

Dey, S. S., Bhatia, R., Sharma, S. R., Parkash, C., and Sureja, A. K. (2013). Effects of chloroplast substituted ogura male sterile cytoplasm on the performance of cauliflower (Brassica oleracea var. botrytis l.) F1 hybrids. *Scientia Horticulturae.* 157, 45–51. doi: 10.1016/j.scienta.2013.04.008

Dieterich, J. H., Braun, H. P., and Schmitz, U. K. (2003). Alloplasmic male sterility in brassica napus (CMS 'Tournefortii-stiewe') is associated with a special gene arrangement around a novel atp9 gene. *Mol. Genet. Genomics* 269 (6), 723–731. doi: 10.1007/s00438-003-0886-3

Dong, X., Kim, W. K., Lim, Y.-P., Kim, Y.-K., and Hur, Y. (2013). Ogura-CMS in Chinese cabbage (Brassica rapa ssp. pekinensis) causes delayed expression of many nuclear genes. *Plant Sci.* 199-200, 7–17. doi: 10.1016/j.plantsci.2012.11.001

Fu, T. (1981). Production and research of rapeseed in the people's republic of China. *Eucarpia Cruciferae Newsl* 6, 6–7.

Geddy, R., Mahé, L., and Brown, G. G. (2005). Cell-specific regulation of a brassica napus CMS-associated gene by a nuclear restorer with related effects on a floral homeotic gene promoter. *Plant J. Cell Mol. Biol.* 41 (3), 333–345. doi: 10.1111/j.1365-313X.2004.02305.x

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47 (W1), W59–W64. doi: 10.1093/nar/gkz238

Heng, S., Liu, S., Xia, C., Tang, H., Xie, F., Fu, T., et al. (2018). Morphological and genetic characterization of a new cytoplasmic male sterility system (oxa CMS) in stem mustard (Brassica juncea). *TAG Theor. Appl. Genet. Theoretische und angewandte Genetik* 131 (1), 59–66. doi: 10.1007/s00122-017-2985-2

Heng, S., Wei, C., Jing, B., Wan, Z., Wen, J., Yi, B., et al. (2014). Comparative analysis of mitochondrial genomes between the hau cytoplasmic male sterility (CMS) line and its iso-nuclear maintainer line in brassica juncea to reveal the origin of the CMS-associated gene orf288. *BMC Genomics* 15 (1), 322. doi: 10.1186/1471-2164-15-322

Hu, Q., Andersen, S., Dixelius, C., and Hansen, L. (2002). Production of fertile intergeneric somatic hybrids between brassica napus and sinapis arvensis for the enrichment of the rapeseed gene pool. *Plant Cell Rep.* 21 (2), 147–152. doi: 10.1007/s00299-002-0491-7

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., et al. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a bloom filter. *Genome Res.* 27 (5), 768–777. doi: 10.1101/gr.214346.116

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., et al. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* 36 (Database issue), D250–D254. doi: 10.1093/nar/gkm796.

Ji, J., Huang, J., Yang, L., Fang, Z., Zhang, Y., Zhuang, M., et al. (2020). Advances in research and application of Male sterility in brassica oleracea. *Horticulturae.* 6 (4), 101. doi: 10.3390/horticulturae6040101

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32 (Database issue), D277–D280. doi: 10.1093/nar/gkh063

Kang, L., Li, P., Wang, A., Ge, X., and Li, Z. (2017). A novel cytoplasmic Male sterility in brassica napus (inap CMS) with carpelloid stamens *via* protoplast fusion with Chinese woad. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00529

Kurusu, T., Koyano, T., Hanamata, S., Kubo, T., Noguchi, Y., Yagi, C., et al. (2014). OsATG7 is required for autophagy-dependent lipid metabolism in rice postmeiotic anther development. *Autophagy.* 10 (5), 878–888. doi: 10.4161/auto.28279

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35 (9), 3100–3108. doi: 10.1093/nar/gkm160

Landgren, M., Zetterstrand, M., Sundberg, E., and Glimelius, K. (1996). Alloplasmic male-sterile brassica lines containing b. tournefortii mitochondria express an ORF 3′ of the atp6 gene and a 32 kDa protein. *Plant Mol. Biol.* 32 (5), 879–890. doi: 10.1007/BF00020485

Lhomme, Y., Stahl, R., Li, X., Hameed, A., and Brown, G. (1997). Brassica nap cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the pol CMS-associated orf224 gene. *Curr. Genet.* 31 (4), 325–335. doi: 10.1007/s002940050212

Li, D. R. (1980). Report on three-lines breeding in brassica napus. *Shaanxi J. Agric. Sci.* 1, 26–29.

Liu, J., Xiang, R., Wang, W., Mei, D., Li, Y., Mason, A. S., et al. (2015). Cytological and molecular analysis of nsa CMS in brassica napus l. *Euphytica.* 206 (2), 279–286. doi: 10.1007/s10681-015-1443-y

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi: 10.1093/nar/25.5.955

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., et al. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45 (5), 573–577. doi: 10.1038/ng.2570

Ogura, H. (1967). "Studies on the new Male-sterility in Japanese radish, with special reference to the utilization of this sterility towerds the practical raising of hybrid seeds," in *Memoirs of the faculty of agriculture,* (Korimoto 1chome 21-24, Kagoshima City, Kagoshima University) 6, 39–78.

Park, H. S., Lee, W. K., Lee, S. C., Lee, H. O., Joh, H. J., Park, J. Y., et al. (2021). Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F(1) hybrid combinations. *Sci. Rep.* 11 (1), 2506. doi: 10.1038/s41598-021-81988-w

Pearson, O. H. (1972). Cytoplasmically inherited male sterility characters and flavor components from the species cross brassica nigra (L) Koch X brassica oleracea l. *J. Am. Soc. Hortic. Science.* 97, 397–402. doi: 10.21273/JASHS.97.3.397

Rawat, D., and Anand, I. (1979). Male Sterility in Indian mustard. *J. Genet. Plant Breed.* 39, 412–415.

Sang, S., Cheng, H., Hao, M., Ding, B., Mei, D., Wang, H., et al. (2021). Mitochondrial localization of ORF346 causes pollen abortion in alloplasmic male sterility. *Crop J.* 9 (6), 1320–1329. doi: 10.1016/j.cj.2021.01.008

Shaya, F., Gaiduk, S., Keren, I., Shevtsov, S., Zemah, H., Belausov, E., et al. (2012). Expression of mitochondrial gene fragments within the tapetum induce male sterility by limiting the biogenesis of the respiratory machinery in transgenic tobacco. *J. Integr. Plant Biol.* 54 (2), 115–130. doi: 10.1111/j.1744-7909.2012.01099.x

Shiga, T., and Baba, S. (1971). Cytoplasmic male sterility in rape plants (Brassica napus l). *Japanese J. Breed.* 21, 16–17.

Singh, M., and Brown, G. G. (1993). Characterization of expression of a mitochondrial gene region associated with the brassica "Polima" CMS: developmental influences. *Curr. Genet.* 24 (4), 316–322. doi: 10.1007/BF00336783

Srivastava, A. P., and Luo, M. (2018). High-resolution cryo-EM analysis of the yeast ATP synthase in a lipid membrane. *Science* 360 (6389), eaas9699. doi: 10.1126/science.aas9699

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30 (12), 2725–2729. doi: 10.1093/molbev/mst197

Tanaka, Y., Tsuda, M., Yasumoto, K., Yamagishi, H., and Terachi, T. (2012). A complete mitochondrial genome sequence of ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (Raphanus sativus l.). *BMC Genomics* 13 (1), 352. doi: 10.1186/1471-2164-13-352

Wang, B., Farooq, Z., Chu, L., Liu, J., Wang, H., Guo, J., et al. (2021). High-generation near-isogenic lines combined with multi-omics to study the mechanism of polima cytoplasmic male sterility. *BMC Plant Biol.* 21 (1), 130. doi: 10.1186/s12870-021-02852-7

Wang, K., Gao, F., Ji, Y., Liu, Y., Dan, Z., Yang, P., et al. (2013). ORFH79 impairs mitochondrial function *via* interaction with a subunit of electron transport chain complex III in honglian cytoplasmic male sterile rice. *New Phytol.* 198 (2), 408–418. doi: 10.1111/nph.12180

Wang, Y., Wang, Q., Hao, W., Li, J., Qi, M., and Zhang, L. (2020). Mitochondrial genome sequencing reveals orf463a may induce Male sterility in NWB cytoplasm of radish. *Genes.* 11 (1), 74. doi: 10.3390/genes11010074

Wan, Z., Jing, B., Tu, J., Ma, C., Shen, J., Yi, B., et al. (2008). Genetic characterization of a new cytoplasmic male sterility system (hau) in brassica juncea and its transfer to b. napus. *TAG Theor. Appl. Genet. Theoretische und angewandte Genetik* 116 (3), 355–362. doi: 10.1007/s00122-007-0673-3

Yamagishi, H., and Bhat, S. R. (2014). Cytoplasmic male sterility in brassicaceae crops. *Breed. Science.* 64 (1), 38–47. doi: 10.1270/jsbbs.64.38

Yamagishi, H., and Terachi, T. (2017). "Cytoplasmic Male sterility and mitochondrial genome variations in radish," in *The radish genome*, vol. p . Eds. T. Nishio and H. Kitashiba (Cham: Springer International Publishing), 93–108.

Yang, Q., Nong, X., Xu, J., Huang, F., Wang, F., Wu, J., et al. (2021). Unraveling the genetic basis of fertility restoration for cytoplasmic Male sterile line WNJ01A originated from brassica juncea in brassica napus. *Front. Plant Sci.* 12, 721980. doi: 10.3389/fpls.2021.721980

Yang, H., Xue, Y., Li, B., Lin, Y., Li, H., Guo, Z., et al. (2022). The chimeric gene atp6c confers cytoplasmic male sterility in maize by impairing the assembly of the mitochondrial ATP synthase complex. *Mol. Plant* 15 (5), 872–886. doi: 10.1016/j.molp.2022.03.002

Yarrow, S. A., Burnett, L. A., Wildeman, R. P., and Kemble, R. J. (1990). The transfer of 'Polima' cytoplasmic male sterility from oilseed rape (Brassica napus) to broccoli (B. oleracea) by protoplast fusion. *Plant Cell Rep.* 9 (4), 185–188. doi: 10.1007/BF00232176

Yu, D., Gu, X., Zhang, S., Dong, S., Miao, H., Gebretsadik, K., et al. (2021). Molecular basis of heterosis and related breeding strategies reveal its importance in vegetable breeding. *Horticulture Res.* 8 (1), 120. doi: 10.1038/s41438-021-00552-9

Zancani, M., Braidot, E., Filippi, A., and Lippe, G. (2020). Structural and functional properties of plant mitochondrial f-ATP synthase. *Mitochondrion.* 53, 178–193. doi: 10.1016/j.mito.2020.06.001

Zhong, X., Chen, D., Cui, J., Li, H., Huang, Y., and Kang, J. (2021). Comparative analysis of the complete mitochondrial genome sequences and anther development cytology between maintainer and ogura-type cytoplasm male-sterile cabbage (B. oleracea var. capitata). *BMC Genomics* 22 (1), 646. doi: 10.1186/s12864-021-07963-x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpls.2022.998082

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Kumar Paritosh,
University of Delhi, India
Yinbo Ma,
Yangzhou University, China

*CORRESPONDENCE
Yang Xiang
xiangyangcell@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 19 July 2022
ACCEPTED 29 September 2022
PUBLISHED 20 October 2022

CITATION
Xie M, Zhao C, Song M, Xiang Y and Tong C (2022) Genome-wide identification and comparative analysis of *CLE* family in rapeseed and its diploid progenitors.
Front. Plant Sci. 13:998082.
doi: 10.3389/fpls.2022.998082

Genome-wide identification and comparative analysis of *CLE* family in rapeseed and its diploid progenitors

Meili Xie[1,2†], Chuanji Zhao[2†], Min Song[1,3†], Yang Xiang[1*] and Chaobo Tong[2]

[1]Guizhou Rapeseed Institute, Guizhou Academy of Agricultural Sciences, Guiyang, China, [2]The Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China, [3]College of Life Science, Qufu Normal University, Qufu, China

Crop genomics and breeding CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE) proteins belong to a small peptide family in plants. During plant development, *CLE* gene family members play a pivotal role in regulating cell-to-cell communication and stem cell maintenance. However, the evolutionary process and functional importance of *CLE*s are unclear in Brassicaceae. In this study, a total of 70 *BnCLE*s were identified in *Brassica napus* (2n = 4x = 38, A_nC_n): 32 from the A_n subgenome, 36 from the C_n subgenome, and 2 from the unanchored subgenome. Meanwhile, 29 *BrCLE* and 32 *BoCLE* genes were explored in *Brassica rapa* (2n = 2x = 20, A_r) and *Brassica oleracea* (2n = 2x = 18, C_o). Phylogenetic analysis revealed that 163 *CLE*s derived from three *Brassica* species and *Arabidopsis thaliana* can be divided into seven subfamilies. Homology and synteny analyses indicated whole-genome triplication (WGT) and segmental duplication may be the major contributors to the expansion of *CLE* family. In addition, RNA-seq and qPCR analysis indicated that 19 and 16 *BnCLE*s were more highly expressed in immature seeds and roots than in other tissues. Some *CLE* gene pairs exhibited different expression patterns in the same tissue, which indicated possible functional divergence. Furthermore, genetic variations and regional association mapping analysis indicated that 12 *BnCLE*s were potential genes for regulating important agronomic traits. This study provided valuable information to understand the molecular evolution and biological function of *CLE*s in *B. napus* and its diploid progenitors, which will be helpful for genetic improvement of high-yield breeding in *B. napus*.

KEYWORDS

CLE peptide, *Brassica napus*, diploid progenitor, orthologous pairs, expression pattern, association mapping analysis, yield traits

Introduction

Peptide hormones are active molecules composed of many to several tens of amino acids and serve as signal molecules to exchange information between cells (Grienenberger and Fletcher, 2015). In plants, there are various polypeptides; CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE), one of the largest families of known polypeptides, is generally 12–13 amino acids in length, including a secretion signal peptide in N-terminus and a highly conserved CLE domain in C-terminus. Domain deletion and domain exchange experiments have indicated that the CLE domain of *CLV3* in *Arabidopsis thaliana* plays an independent role in adjacent flanking sequences (Fletcher et al., 1999; Rojo et al., 2002). With the help of matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry, an active 12-amino acid CLV3 polypeptide molecule (corresponding to the CLE domain, with one amino acid removed from each side), is isolated from *A. thaliana* over-expressing *CLV3*, and the peptide synthesized *in vitro* is functional (Tatsuhiko, 2006). Exogenous application of synthetic CLV3, CLE19, and CLE40 polypeptides exhibits a similar phenotype with overexpression of these *CLE* genes. CLV3 polypeptides also restore the *clv3-2* mutant phenotype (Fiers et al., 2006). These results indicate that CLE polypeptides are the active form of CLE family proteins.

It is difficult to clearly identify the function of each CLE peptide owing to its small size and high sequence conservation. However, the application of new technologies such as gene editing makes it easier to study their functions. However, it is easier to study their functions based on the application of new technology, like gene editing. Now, many results show that CLE polypeptides play important roles in plant development and hormone and stress response. In shoot apical meristems (SAMs), the stem cell homeostasis is maintained by a dynamic negative feedback loop involved in the CLV3-WUSCHEL (WUS) pathway. The transcription factor gene *WUS*, which interacts with SHOOT MERISTEMLESS (STM), can promote the expression of CLV3; meanwhile, overmuch CLV3 will suppress the expression of STM and WUS to maintain the stem cell population in the SAM (Schoof et al., 2000). In addition to participating in the division and differentiation of meristem cells, CLE polypeptides also play a key role in the development of seeds (Fiume, 2010). With the use of the method of promoter fusion GUS, *CLE8* and *WUSCHEL-related homeobox 8* (*WOX8*) are expressed in the embryo and endosperm during the early stages of seed development in *Arabidopsis* (Fiume and Fletcher, 2012). The number of embryo and endosperm cells decreases in the *cle8-1* mutant, suggesting that *CLE8* promotes the proliferation of embryo and endosperm cells. The length and width of seeds produced by *CLE8*-overexpressing plants and *wox8-1* mutant seeds significantly increase and decrease, respectively, compared with those of wild-type seeds. *WOX8* expression significantly increases in CLE8-overexpression lines, indicating that *CLE8* promotes its expression (Fiume and Fletcher, 2012; Song et al., 2013). Some CLE peptides likely interact with hormones. For example, CLE26 expressed at the phloem pole, regulated root architecture, and its expression is significantly enhanced by auxin treatment. CLE26 can affect the activity of the polar auxin transporter in the auxin signaling (Czyzewicz et al., 2015). The shoot growth is related to the root-expressed CLE6 under the gibberellin effect (Bidadi et al., 2014). In addition to development and hormone response, CLE also mediates responses to various abiotic stress. The expression of CLE25 is increased when the root is hydropenic, and then the root-derived CLE25 peptide moves into the leaves to modulate the closure of the stoma (Takahashi et al., 2018). In response to low-sulfate conditions, the expression of CLE2 and CLE3 in roots is reduced, which could inhibit lateral root development (Dong et al., 2019).

CLE genes are widely present in plants and even in several plant-parasitic nematodes. Three *CLE*s expressed in the maize endosperm (*Esr1*, *Esr2*, and *Esr3*) are supposed to be involved in signal transduction between embryo and endosperm during early development (Bonello et al., 2000). *ZmCLE7* and *ZmFCP1* are CLV3 homologs, and promoter editing performed by CRISPR-Cas9 increases many grain-yield-related traits in maize (Liu et al., 2021). Two *CLE*s from *Lotus japonicus* (*LjCLE-RS1* and *LjCLE-RS2*) are repressors of excess root nodulation (Okamoto et al., 2013). An *AtCLE19*-like gene with high expression in the flower bud, pistil, and embryo of *Brassica napus* was ectopically expressed in *A. thaliana*, resulting in large heads (Fiers et al., 2004). The mutations of *BnCLV3*s induced by CRISPR/Cas9 could result in multilocular siliques and an increase in seed production (Yang et al., 2018). Similarly, *BrCLV3* mutations conferred multicellular pods in *Brassica rapa* (Fan et al., 2014). In *Raphanus sativus*, overexpression of *RsCLE2* and *RsCLE19* increased the number of xylem elements (Gancheva et al., 2016). Moreover, *CLE*s, which may assist the infection process, are identified in cyst nematodes (Wang J. et al., 2011). Current research suggests that the *CLE*s are highly conserved and involved in biological evolution, especially in plants.

Recent genome-wide analyses have identified CLE genes in many plants (Zhang et al., 2014; Hastwell et al., 2015; Gancheva et al., 2016; Han et al., 2016; Han et al., 2020). The allotetraploid species *B. napus* (2n = 4x = 38, A_nC_n), an important oil crop, was formed from the hybridization between *B. rapa* and *Brassica oleracea* (2n = 2x = 18, C_o) at about 7,500 years ago; therefore, the phylogenetic relationship of *Brassica* provides a good basis for studying the evolution of gene family (Allender and King, 2010). Based on bioinformatics and comparative genomic approaches, this study performed a multidimensional investigation for *CLE*s in rapeseed and its diploid progenitors, including genome-wide identification, molecular characterization, phylogenetic analysis,

synteny analysis, expression profiling in different tissues, and regional association mapping analysis. The results will provide useful information for biological functions and molecular evolution in *Brassica*, which also supply candidate genes for genetic improvement in rapeseed breeding.

Materials and methods

Identification of *CLE*s in Brassicaceae

Multiple TBLASTN and BLASTN searches were performed for CLEs identification in *B. napus* reference genome (http://brassicadb.org/brad/downloadOverview.php) (Wang et al., 2015) based on *A. thaliana CLE* [expected threshold (e−10)]. The results were then validated by the Conserved Domain Database (CDD) in the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/cdd) (Lu et al., 2020) and Modular Architecture Research Tool SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2020) to confirm the authenticity of the CLE domain in the open reading frame. Open reading frames of homologous chromosome regions were confirmed for potential unannotated or truncated duplicates of CLEs. *BrCLE*s in *B. rapa* and *BoCLE*s in *B. oleracea* (http://brassicadb.org/brad/downloadOverview.php) were obtained as described above to explore the evolution of *CLE*s in *Brassica*.

Gene structure, conserved motifs, and *cis*-acting regulatory elements analysis

Gene structures and conserved motifs for the *CLE*s were constructed *via* TBtools. Logo diagrams used to define consensus sequences were obtained using multiple sequence alignments for each BnCLE peptide group (I–V), including *A. thaliana*, *B. rapa*, and *B. oleracea* by TEXshade (Beitz, 2000). The signal peptides of CLEs were identified *via* SMART (Letunic et al., 2020). PlantCARE was carried out to predict the promoters in the 2-kb region before the start codon for the *cis*-element identification (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Magali, 2002). Gene Structure Display Server (GSDS 2.0) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to exhibit the gene structure (Hu et al., 2015), and the heatmap was visualized by R package (https://cran.r-project.org/).

CLE gene duplication pattern and synteny analysis

In order to investigate the synteny relationship of CLEs in Brassicaceae, all the *CLE*s were searched as "syntenic genes" in Brassicaceae Database (BRAD) (http://brassicadb.cn/#/). At the same time, TBtools (Chen et al., 2020) was used to detect gene duplication patterns and verify the synteny relationship, as well as calculate the ratio of non-synonymous to synonymous substitutions (Ka/Ks) of syntenic gene pairs. Orthologous *CLE*s located on syntenic chromosome blocks were displayed by Circos software (Krzywinski et al., 2009).

Phylogenetic analysis of CLEs

Multiple alignments of the CLE peptide sequences from the four Brassicaceae species were performed by the ClustalW (Larkin, 2007). Phylogenetic analysis was generated using the MEGA7 soft with the maximum likelihood (ML) method, 1,000 bootstrap replications, and the JTT+G model (Kumar et al., 2016). The tree was visualized using Evolview (https://www.evolgenius.info/evolview/) (He et al., 2016).

Prediction of protein–protein interactions

Protein–protein interactions in *A. thaliana* were obtained in the STRING database (https://www.string-db.org/). The analysis and demonstration of CLEs interactions in *B. napus* followed the description based on the previous study (Xie et al., 2022).

Transcriptome expression pattern of *CLE*s

RNA-seq raw data of siliques, leaves, flowers, and stems for *B. napus*, *B. rapa*, and *B. oleracea* were downloaded from NCBI (ProjectID: PRJNA489323); after being filtered with Trimmomatic (Bolger et al., 2014), the clean data without adapters and low-quality bases were aligned with the reference genome (http://brassicadb.org/brad/downloadOverview.php) using hisat2 (Kim et al., 2015). Based on the mapping results, the gene expressions (fragments per kilobase million (FPKM)) were counted by Stringtie (Pertea et al., 2015), and the heatmaps were drawn by R.

Quantitative real-time PCR analysis

The *B. napus* ZS11 was used in this study. Tissues like flowers, roots, stems, leaves, immature pods, immature seeds, and apical meristems were collected and extracted using an RNeasy Extraction Kit (Invitrogen, Carlsbad, CA, USA). Quantitative real-time PCR (qRT-PCR) was carried out by referring to a formerly described protocol (Zhao et al., 2021). The expression level was displayed by heatmap in R. Expression

patterns of five genes were shown by TBtools-eFP (Chen et al., 2020) (http://yanglab.hzau.edu.cn/BnTIR/eFP).

Genetic variation of *CLE*s in *Brassica napus* core accessions

To investigate the genetic variation of *CLE* genes in *B. napus*, a panel of 204 rapeseed accessions was selected for this work (Zhao et al., 2022). The single-nucleotide polymorphism (SNP) information of *CLE* genes was retained from the previous study. After annotation with SnpEff (Cingolani et al., 2012), the distribution was analyzed to inspect the position of variations. The agronomic traits for three consecutive years (2014–2016) including plant height, branch number, initial branch height, length, silique number, and silique density of main inflorescence, main inflorescence, silique length, seed number per silique, main inflorescence seed density, thousand seed weight, seed weight per silique, and seed weight of the main inflorescence were surveyed and handled with the best linear unbiased prediction (BLUP). To study the potential impact of *CLE* genes on agronomic traits in *B. napus*, SNPs within 30 kb upstream and downstream of *CLE*s were used for the investigation of its potential impact. Regional association analysis considering population structure and relative kinship was conducted by EMMAX (Kang et al., 2010).

Result

Identification of CLEs in *Brassica napus* (Bn), *Brassica rapa* (Br), and *Brassica oleracea* (Bo)

A genome-wide analysis of *CLE* genes in *B. napus* and its diploid progenitors was performed involving multiple BLAST queries and iterative queries, followed by domain validation and removal of false positives (i.e., no CLE domain). Finally, a total of 29 (BrCLEs), 32 (BoCLEs), and 70 (BnCLEs) genes were identified (Table S2). The total number of *BrCLE*s and *BoCLE*s in the two diploid progenitors was lower than that of *BnCLE*s in the allotetraploid rapeseed, indicating that CLE gene expansion event has occurred in *B. napus* during polyploidization.

The *CLE*s in the three species of *Brassica* species were renamed according to the AtCLEs in *Arabidopsis* based on the naming conventions for the *B. genus* (Ostergaard and King, 2008), with the last letter "a" indicating the highest homology with *Arabidopsis*, and then "b", and so on. In *B. napus*, the capital letters A and C following "*Bn*" represented the A_n and C_n subgenomes, respectively.

Phylogenetic analysis of *CLE*s in Brassicaceae

To explore the evolution of the *CLE*s in *Brassica*, we constructed a phylogenetic tree using full-length amino acid sequences from Brassicaceae (Figure 1). The results showed *CLE*s clearly divided into seven subfamilies (I to VII). Subfamily V was the largest (42 members), followed by subfamily I (38 members). There were 17, 11, 23, 17, and 14 members in subfamilies II, III, IV, VI, and VII, respectively. The *CLE*s of the four species were distributed in all subfamilies. Most *AtCLE*s in each subfamily matched multiple sets of orthologs from *B. napus* and its two progenitors. A sister pair demonstrated the closest genetic relationship in a phylogenetic tree. A total of 56 sister pairs were observed. The majority of the sister pairs were orthologous gene pairs between the A_n (or C_n) subgenomes of *B. napus* and *B. rapa* (or *B. oleracea*), with 18 A_n–A_r pairs and 19 C_n–C_o pairs. These results supported that the gene duplication events happened in the *B. napus* genome and indicated that *CLE* orthologous genes of distinct subfamilies were highly conserved in the respective genome.

Duplication pattern and chromosome localization analysis of *CLE*s

The chromosomal locations of *BnCLE*s, *BrCLE*s, and *BoCLE*s were investigated according to their physical positions (Figure 2). The *BnCLE*s were asymmetrically distributed on the 19 chromosomes in *B. napus*. There were a total of 32 *BnCLE*s in the A_n subgenome and 36 in the C_n subgenome, which were similar to those in *B. rapa* (A_r, 29) and *B. oleracea* (C_o, 32). The remaining two *BnCLE* genes were located on unanchored scaffolds (Table S2). Each chromosome harbored at least one *CLE* gene. Chromosome A_n07 in *B. napus* carried the most *CLE*s (seven *CLE*s). On chromosomes A_n01, C_n01, and C_n08, only one *CLE* gene was found. Furthermore, many *CLE*s retained their relative position in A_r and A_n, whereas only a portion of *CLE*s retained their relative position in C_o and C_n. For example, the same number of *CLE*s in Ar01–An01, Ar05–An05, Ar06–An06, and Ar08–An08 was observed, which also showed similar locations in the chromosomes (Figure 2).

We searched four duplicated types in each Brassicaceae species, including dispersed, proximal, tandem, and whole-genome duplication (WGD) (Table S3). We found that 56 of 70 *BnCLE*s were derived from segmental duplication/whole-genome triplication (WGT). Therefore, it appeared that segmental duplication/WGT played an important role in the *BnCLE* expansion. In addition, we examined *CLE* gene expansion patterns in *B. oleracea* and *B. rapa*, finding that

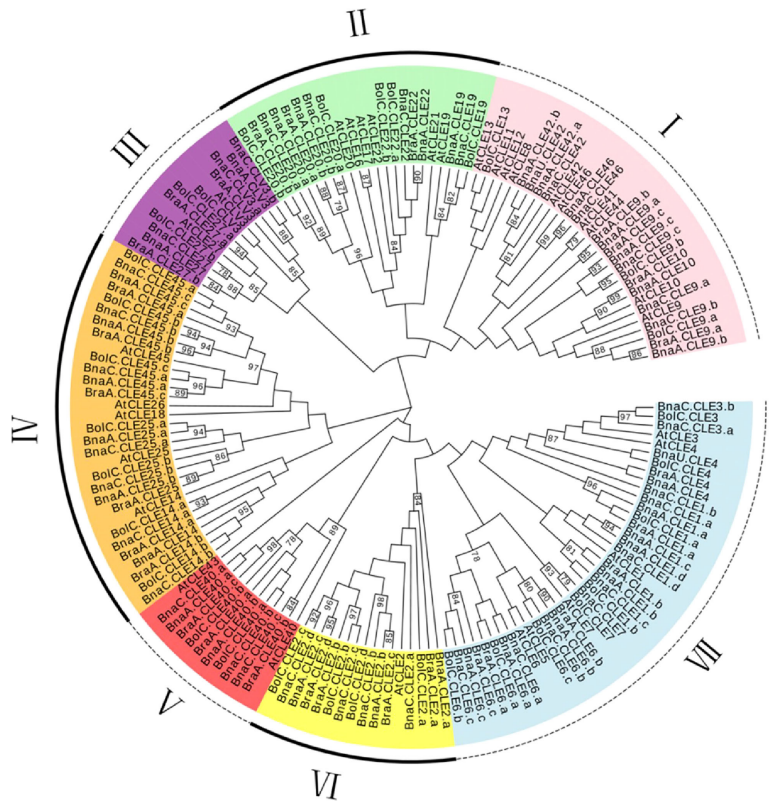


FIGURE 1
Phylogenetic analysis of *Brassica napus*, *Arabidopsis thaliana*, *Brassica rapa*, and *Brassica oleracea CLE* genes. This unrooted phylogenetic tree was constructed by MEGA 7.0 using the maximum likelihood method with 1,000 bootstrap replicates. Only values above 50% were displayed.

most *BoCLE*s (53.1%; 17/32) and *BrCLE*s (89.7%; 26/29) were also derived from segmental duplication/WGT (Table S3). One cluster of *CLE* tandem repeat genes was only present in *B. rapa* and *B. oleracea*, but not in *B. napus*.

Structure analysis of the *CLE*s

A great amount of the identified *CLE*s lacked predicted introns, except for 20 genes, including one to three introns (Table S2, Figure S1). Like other plants, the multiple protein sequence alignment of CLE pre-propeptides in *B. napus* contained a signal peptide at the N-terminus, a central variable domain, and a CLE domain at the C-terminus. The signal peptide sequences and the variable domain among distinct genes were not highly conserved, except between orthologous genes. The highly conserved CLE domain that contained approximately 12/13 aa and residues at different positions had different levels of conservation. To study the sequence of the conserved CLE domains and the degree of their conservation in different Brassicaceae species, multiple sequence alignment was used to generate the protein sequence logos in *B. rapa*, *B. oleracea*, *B. napus*, and *Arabidopsis* (Figure 3). The results showed that conserved residues patterns of the CLE domain were remarkably similar in these four plants (Figure 3). The conservation of residues 1, 4, 6, 8, 9, 11, and 12 suggested that they may be critical to the function of CLE mature peptides. These results indicated that *CLE* gene family was relatively conserved, while some motif sequences changed slightly during *Brassica* evolution, which possibly contributed to extended special biological function.

Cis-elements in the promoters can affect gene expression (Davuluri et al., 2003; Klaas, 2009). Therefore, *CLE* gene promoters of these three species were investigated using PlantCARE (Magali, 2002). Four to 12 *cis*-elements involving development, hormone, and stress were identified in each *BnCLE* promoter, while 2–12 and 4–13 were found in *BrCLE*s and *BoCLE*s, respectively (Figure S2). ARE has the most elements in these three species and existed in 87.1% (61/70) *BnCLE* promoters, 93.1% (27/29) *BrCLE*s, and 96.9% (31/32) *BoCLE*s; this element is essential for anaerobic induction (Figure 4). Hormone-responsive elements involved in abscisic acid responsiveness, methyl jasmonate responsiveness, and ethylene responsiveness were also very common in *CLE*

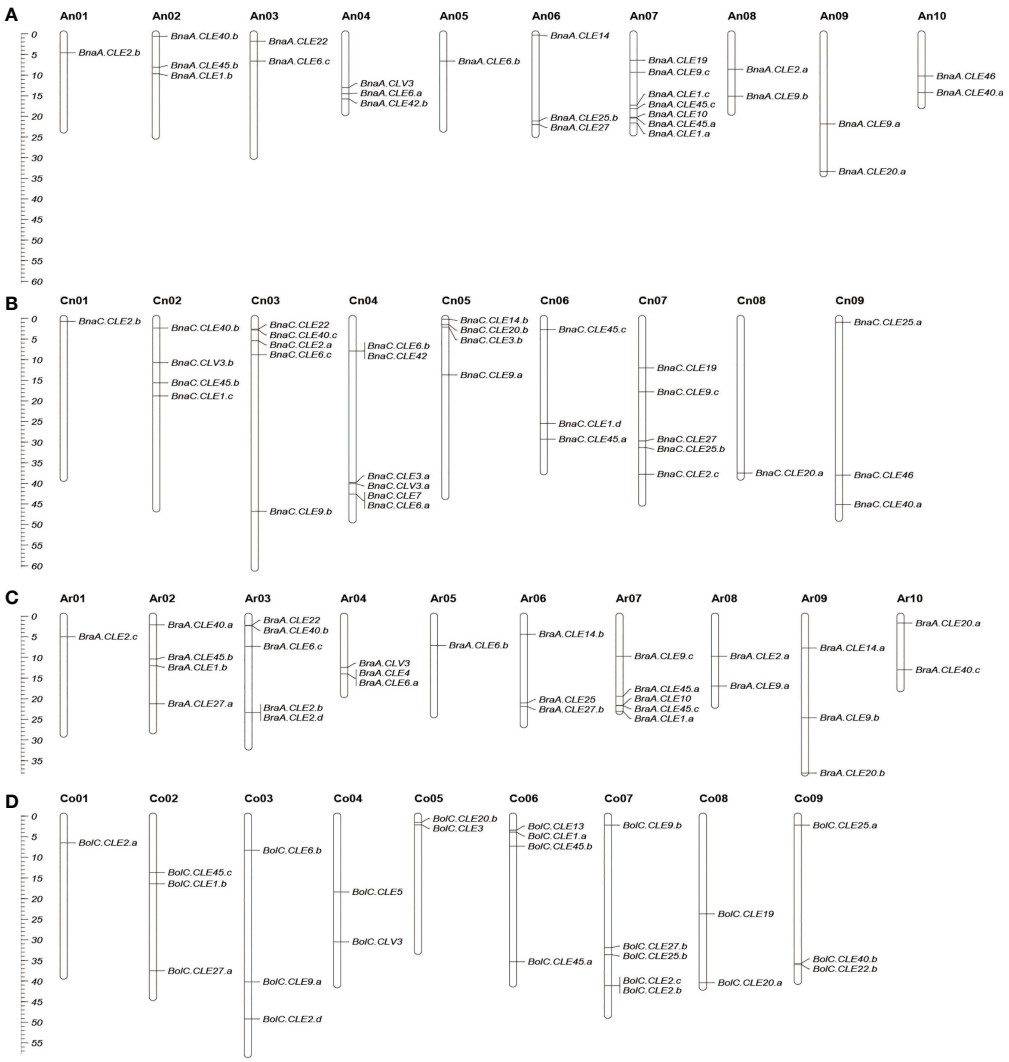


FIGURE 2
Chromosomal location of *CLE*s in *Brassica napus* **(A, B)**, *Brassica rapa* **(C)**, and *Brassica oleracea* **(D)**. Partial *CLE*s in *B. oleracea* and *B. napus* located in unassembled scaffolds were not shown. The scale on the left is in megabases (Mbs).

promoters: 80% (56/70), 86.2% (25/29), and 87.5% (28/32) of *BnCLE*, *BrCLE*, and *BoCLE* promoters included ABRE, respectively. In development elements, GCN4 motif involved in endosperm expression (25.7%, 18/70 in *BnCLE*s; 37.9%, 11/29 in *BrCLE*s; and 25%, 8/32 in *BoCLE*s), O2 site involved in zein metabolism regulation (40%, 28/70 in *BnCLE*s; 31%, 9/29 in *BrCLE*s; and 34.4%, 11/32 in *BoCLE*s), CAT-box related to meristem expression (27.1%, 19/70 in *BnCLE*s; 31%, 9/29 in *BrCLE*s; and 37.5%, 12/32 in *BoCLE*s) and circadian (32.9%, 23/70 in *BnCLE*s; 24.1%, 7/29 in *BrCLE*s; and 37.5%, 12/32 in *BoCLE*s) were common. According to the cluster results, some similar *CLE* promoters had similar *cis*-elements, like *CLE4*, *CLE19*, and *CLE42*, while other *CLE*s were clustered into different groups.

Orthologous relationship and synteny analysis of *CLE* genes

Syntenic genes are orthologous genes located in syntenic fragments between different species that derive from a shared ancestor. We obtained the syntenic genes of *CLE* genes of *Arabidopsis* in three *Brassica* species by searching "syntenic gene" in BRAD (Wang et al., 2015) and showed these collinearity relationships using Circos (Krzywinski et al., 2009) software between the A_n and C_n subgenomes of *B. napus* and its two diploid progenitors (Figure 5, Table S4). A total of 101 *CLE* genes in three *Brassica* species showed conserved synteny with those in *A. thaliana* and were positioned in the same conserved chromosomal blocks, such as A, B, D, E, F, I, J, L, R, U, and Wb

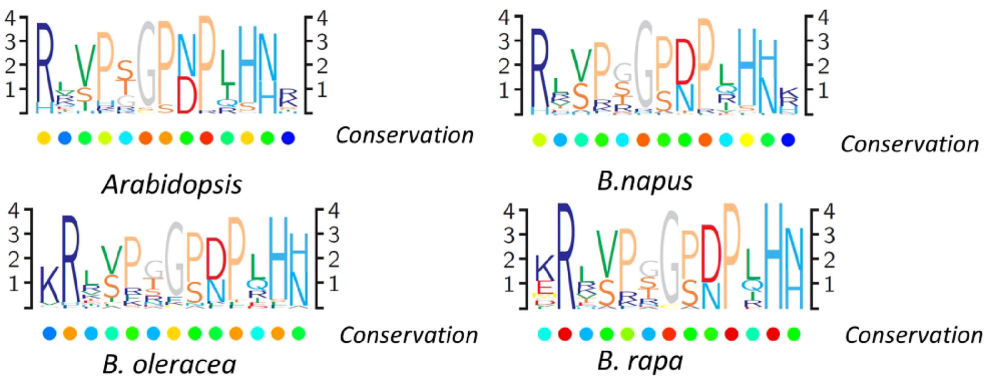


FIGURE 3
CLAVATA3/EMBRYO SURROUNDING (CLE) domain consensus sequences from *Brassica napus*, *Brassica rapa*, *Brassica oleracea*, and *Arabidopsis* pre-propeptides. Logo diagrams illustrate the 13 amino acid CLE domain consensus sequences, determined from multiple sequence alignments generated for each species.

(Schranz et al., 2006). In addition, syntenic genes in three *Brassica* species were divided into three fractionated subgenomes, which were specified as LF (least-fractionated), MF1 (medium-fractionated), and MF2 (most-fractionated) according to the extent of gene retention (Chalhoub et al., 2014). There were 41, 31, and 27 *CLE* genes caught in the LF, MF1, and MF2 subgenomes, respectively (Table S4). A total of 27 *AtCLE* genes retained corresponding syntenic genes in the three *Brassica* species. The existing forms of syntenic genes in the genomes of three *Brassica* species were different. The first type was that syntenic genes of *AtCLE* were completely preserved in the same block of synteny in the A_r, C_o, A_n, and C_n subgenomes, such as *AtCLV3*. The second type was that *AtCLE* genes were retained in the A_r or/and C_o genome but lost in *B. napus* genomes, such as *AtCLE4*, *AtCLE5*, *AtCLE6*, and *AtCLE14*. The third type was that *AtCLE* genes were retained in *B. napus* genome but lost in *B. rapa* or *B. oleracea* genomes, such as *AtCLE1*, *AtCLE9*, and *AtCLE40*. The results showed that the expansion of *CLE* gene family was also accompanied by gene loss.

To comprehend whether natural selection acted on the evolution of *CLE* gene family in *B. napus*, selection pressure analysis was performed on the syntenic *CLE* gene pairs between A_n and A_r, C_n, and C_o. The non-synonymous rate (Ka) and synonymous rate (Ks) values were calculated. The Ka/Ks ratio > 1 represents positive selection, the Ka/Ks ratio = 1 represents neutral selection, and the Ka/Ks ratio < 1 represents purifying selection (Nekrutenko, 2002). The Ka/Ks ratios of the syntenic gene pairs are shown in Table S5. The Ka/Ks ratios for some *CLE* syntenic gene pairs were >1, such as *BnaA.CLE.1c* and *BnaC.CLE.1d*, *BnaC.CLE.2b* and *BolC.CLE.2b*, *BnaA.CLE.10* and *BraA.CLE.10*, and *BnaA.CLE.27* and *BnaC.CLE.27*, which indicated that these genes were subject to positive selection pressure. Many syntenic gene pairs had no Ka/Ks value in *B. napus* because these two genes had the same sequence or large differences. The rest were less than one, which indicated that they underwent purifying selection during the evolution process and may preferentially perform conserved functions.

Predicted protein interactions of BnCLEs, BrCLEs, and BoCLEs

As polypeptide hormones, CLEs need to combine with receptor proteins to transmit signals between cells. To investigate the involved biological process in rapeseed and its diploid progenitors, protein–protein interaction networks were predicted based on known protein interactions in *Arabidopsis*. A total of 1,061 proteins *Arabidopsis* proteins interacted with CLEs, resulting in 3,997 proteins in rapeseed (Figure 6A). CLEs interacted with other proteins but also interacted with each other. Taking out the interacted genes for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis indicated that they played important roles in zeatin biosynthesis, amino acid metabolism (like tryptophan, arginine, and proline), polysaccharide biosynthesis, and ion channels (Figure 6B). Gene Ontology (GO) enrichment analysis showed that their molecular functions mainly were receptor serine/threonine kinase binding and phosphorelay response regulator activity; their biological processes were transmembrane receptor protein tyrosine kinase signaling pathway, polarity specification of adaxial/abaxial axis, stamen development, and lateral root development (Figure 6C). As to *CLE* genes in *B. rapa* and *B. oleracea*, 1,334 and 1,356 proteins were identified as their interacted proteins, respectively. The Gene Ontology enrichment analysis (Figures S3, S4) indicated that these proteins were important in the pattern specification process, shoot system morphogenesis, meristem maintenance, root morphogenesis, and leaf development.

10.3389/fpls.2022.998082

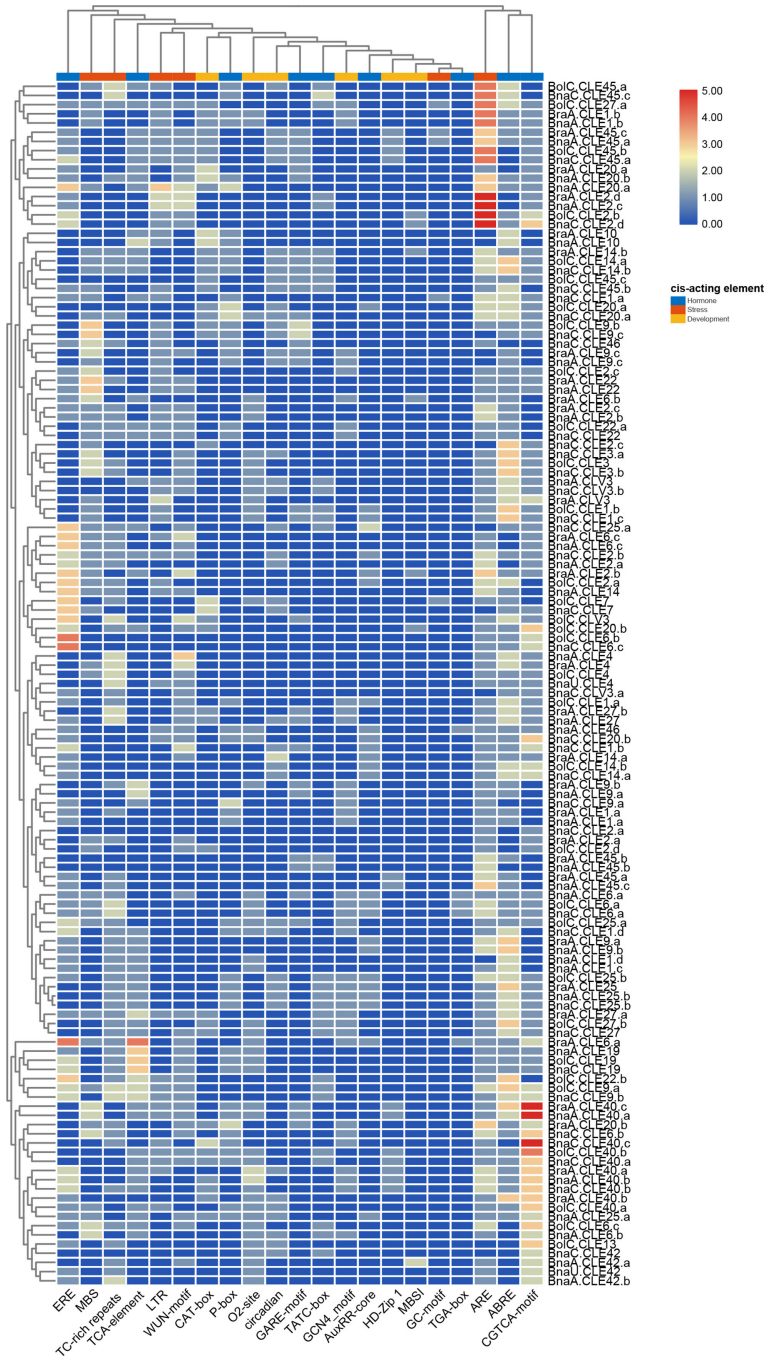


FIGURE 4
Cis-acting elements analysis of *CLE*s in *Brassica napus*, *Brassica rapa*, and *Brassica oleracea*. The color scale bar indicates the numbers of *cis*-acting elements in the promoters.

Expression patterns of *CLE* genes in various tissues

According to the RNA-seq data in siliques, leaves, flowers, and stems of *B. napus*, *B. rapa*, and *B. oleracea*, the expression patterns of *CLE*s are illustrated in Figures S5-S7. The majority of *CLE*s showed low or no expression in all four tissues. Among the expressed *CLE*s, more *CLE* genes were highly expressed in stems in three *Brassica* species, like *BnaC.CLE42*, *BnaA.CLE46*, *BnaC.CLE46*, *BolC.CLE20.b*, *BraA.CLE6.b*, and *BraA.CLE6.c*,

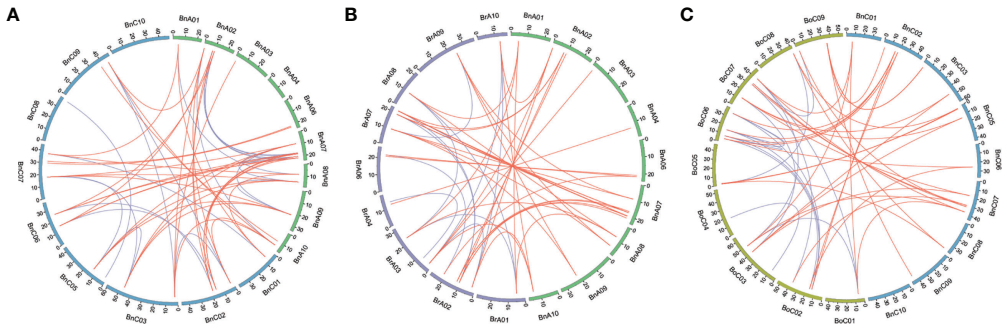


FIGURE 5
Genome-wide synteny analysis for *CLE*s among *Brassica napus*, *Brassica rapa*, and *Brassica oleracea*. **(A)** Synteny analysis of *CLE*s on An and Cn subgenomes in *B. napus*. **(B)** Synteny analysis of *CLE*s between A_n subgenome of *B. napus* and *B. rapa*. **(C)** Synteny analysis of *CLE* genes between Cn subgenome of *B. napus* and *B. oleracea*. Inside the circos, brown lines linked the syntenic orthologs, and blue lines linked the syntenic paralogs.

whereas *BnaA.CLE9.b*, *BolC.CLE27.a*, *BolC.CLE14.b*, and *BraA.CLE14.a* were highly expressed in siliques. Only a few paralog *CLE* genes had a similar expression pattern, like *BraA.CLE27.a/BraA.CLE27.b*, *BnaC.CLE42/BnaA.CLE42.b/BnaA.CLE42.a/BnaU.CLE42*, *BolC.CLE9.a/BolC.CLE9.b*, and *BolC.CLE22.a/BolC.CLE22.b*. In three *Brassica* species, orthologs like *BnaA.CLE20.b/BraA.CLE20.b/BolC.CLE20.b* and *BnaC.CLV3.a/BraA.CLV3/BolC.CLV3* displayed a similar expression pattern. Meanwhile, orthologs in *B. rapa* and *B. oleracea* showed different expression patterns; for example, *BraA.CLE4* was mainly expressed in leaves, whereas *BolC.CLE4* was not. These results suggest that some *CLE* genes have shown functional divergence during their evolution.

Furthermore, we performed a qPCR analysis in *B. napus* in various tissues—flowers, roots, stems, leaves, immature pods, immature seeds, and apical meristems. Similar to the transcriptome data, *BnCLE* genes had significantly different expression patterns in different tissues (Figure S8). *BnaA.CLE1.b*, *BnaA.CLE2.b*, *BnaC.CLE2.a/b/c*, *BnaA.CLE14*, and *BnaC.CLE14.b* were more highly expressed in roots than in the other tissues examined (Figure 7A), whereas *BnaA.CLE6.a/b*, *BnaC.CLE6.b*, *BnaC.CLE7*, *BnaA.CLE4*, *BnaU.CLE4*, *BnaA.CLE1.c*, *BnaC.CLE1.a/c*, *BnaA.CLE9.a/b/c*, and *BnaC.CLE9.a/b/c* were more highly expressed in immature seeds than in the other tissues (Figure 7B). *BnaA.CLE45.c*, *BnaA.CLE19*, *BnaC.CLE19*, and *BnaC.CLV3.a/b* were more highly expressed in apical meristems than in the other tissues (Figure 7C). *BnaC.CLE2.d* and *BnaC.CLE46* were expressed at high levels in root and stem (Figure 7D), while *BnaA.CLE1.a/b*, *BnaA.CLE2.a*, *BnaC.CLE6.a*, and *BnaC.CLE14.a* were expressed at high levels in root and immature seeds (Figure 7E). The materials and sampling time of the qPCR experiments in this study were different from the data downloaded from NCBI; therefore, the results were not completely consistent. However, there was some similarity: *BnCLE*s showed different expression patterns in different tissues and were modestly expressed in leaves. The expression patterns of most *BnCLE*s pairs were significantly different, suggesting that their roles were dissimilar. For example, *BnaA.CLE1.b* was highly expressed in the roots, whereas *BnaC.CLE1.c* was more highly expressed in immature seeds. *BnaA.CLE1.c* was highly expressed in immature seeds, whereas *BnaA.CLE1.d* was similarly expressed in the different tissues. *BnaA.CLE22* was highly expressed in the stem, whereas *BnaC.CLE22* was more highly expressed in apical meristems. *BnaA.CLE46* was highly expressed in the stem, whereas *BnaC.CLE46* was the more highly expressed root, followed by stem.

Regional association mapping for *BnCLE*s

A natural population including 204 accessions was selected to analyze the genetic variations of *CLE*s, and a total of 115 SNPs were identified; on average, 1.6 SNPs were detected for each *CLE* gene, which was apparently less than the whole genome level (28.9 SNPs per gene). The total and average number of SNPs in the A subgenome (74 SNPs, 2.2 SNPs/gene) were higher than in the C subgenome (41 SNPs, 1.1 SNPs/gene). Considering *CLE* gene length, the SNP density of the A subgenome was 5.7 SNPs/kb, whereas it was 2.2 SNPs/kb in the C subgenome. Moreover, the SNP distribution among paralogous genes was also unequal; like for BnaA.CLE25.a/BnaA.CLE25.b/BnaC.CLE25.a/BnaC.CLE25.b, the SNP numbers were 9/1/14/0. Finally, SNP annotation implied that 73 SNPs were in exon regions and 37 SNPs could lead to missense mutations.

In this study, 12 agronomic traits (Figure S9) and SNPs within the range of 30 kb upstream and downstream of *BnCLE*s were used to study the impact of *BnCLE*s in *B. napus*. A total of 12 *CLE*s were significantly associated with at least one agronomic trait

10.3389/fpls.2022.998082

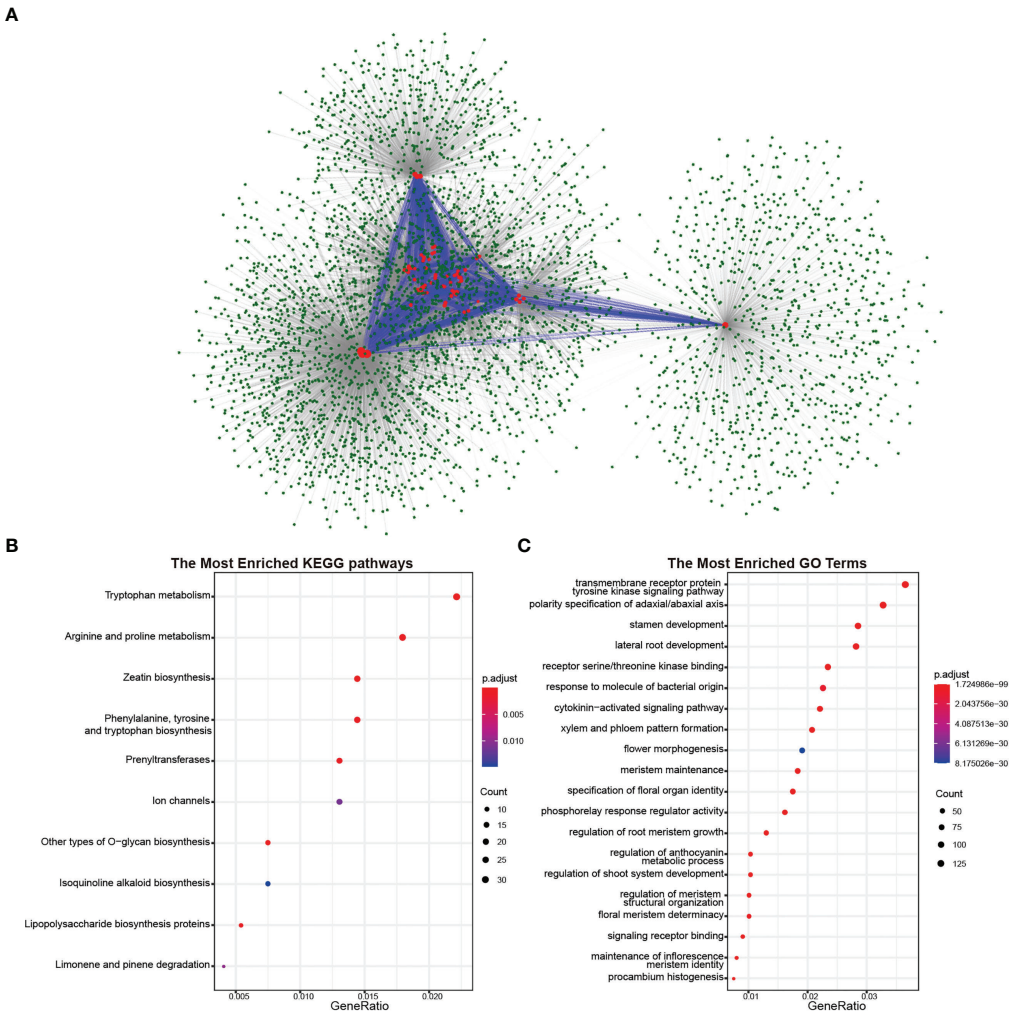


FIGURE 6
Proteins interacted with CLE proteins in Brassica napus. **(A)** Protein–protein interaction network of CLE proteins in B. napus. The red circles represent the CLE proteins, and the green circles represent proteins interacting with CLE proteins. The blue lines represent the interaction between CLE proteins, and the gray lines represent the interaction between CLE proteins and other proteins. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of proteins interacted with CLE proteins. **(C)** Gene Ontology enrichment analysis of proteins interacted with CLE proteins.

($p < 0.0001$) (Table S6), including plant height, branch height, main inflorescence length, seed weight per silique, silique number, silique density, seed density, and seed weight of the main inflorescence. For example, there was no variation in the gene sequence of *BnaC.CLV3.a*, but SNPs in the upstream region (6 kb) were significantly associated with seed weight per silique; the population was clearly divided into two haplotypes based on the SNPs, and the t-test displayed the significant difference in seed weight per silique between these two groups (Figure 8A). According to JASPAR, the SNP region (TCCGTACA) was predicted as the binding site of C2H2 zinc finger factors (SPL3). Moreover, there were another four genes also strongly associated with yield traits, like *BnaA.CLE20.b* and *BnaC.CLE20.b*; SNPs in the upstream region (1.4 and 3.4 kb) were associated with main inflorescence seed density (Figures 8B, C). Their interacted proteins were not only enriched in receptor serine/threonine kinase binding (GO:0033612), signaling receptor binding (GO:0005102), and cell–cell signaling involved in cell fate commitment (GO:0045168) but also enriched in GO terms including regulation of meristem structural organization (GO:0009934), maintenance of meristem identity (GO:0010074), and maintenance of root meristem identity (GO:0010078). Therefore, it was speculated that these *CLE*s could affect the development of rapeseed and result in phenotypic variations.

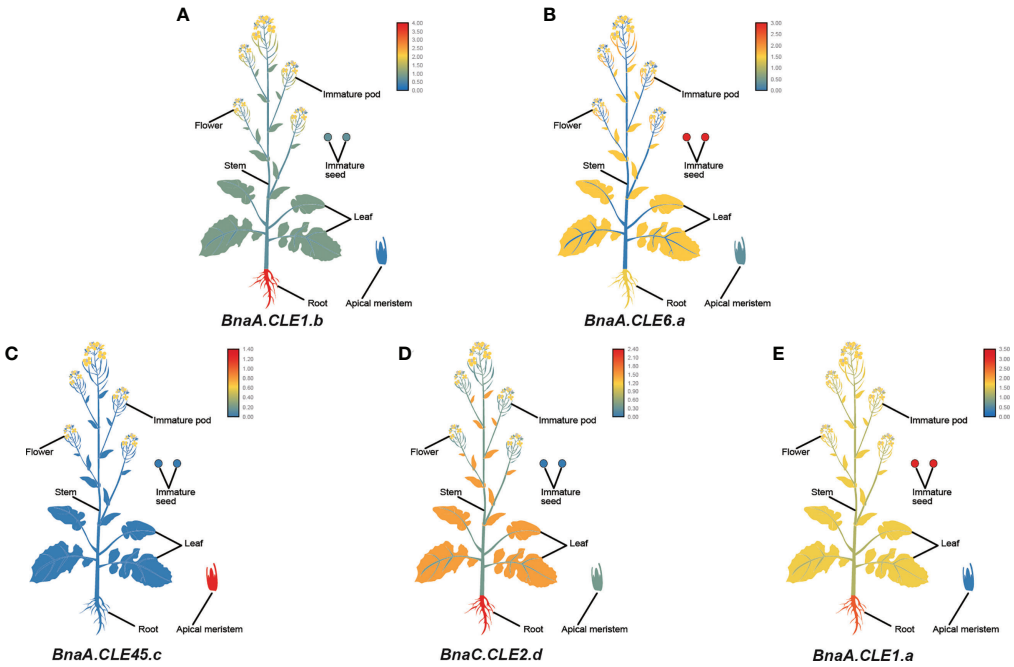


FIGURE 7
The expression patterns of five selected CLE genes **(A–E)** in Brassica napus plants. Expression data were processed with log10 normalization. The color scale represents relative expression levels from low (blue color) to high (red color).

Discussion

CLE peptides are plant-specific peptide hormones that act as mediators of cell-to-cell communication (Fletcher, 2020). Genome-wide studies on *CLE* gene family have been performed in tomato, soybean, wheat, and populus (Zhang et al., 2014; Hastwell et al., 2015; Han et al., 2016; Li Z. et al., 2019; Han et al., 2020). In this study, systematic identification, classification, evolution, expression, and association mapping analysis were performed in *B. napus*. In total, we identified 70, 32, and 29 *CLE*s in *B. napus*, *B. oleracea*, and *B. rapa*, respectively, which were less than those in the previous work (Han et al., 2020), and the possible reason was more strict parameters used in this work. These *CLE*s were divided into seven subfamilies according to the phylogenetic tree. Furthermore, this work analyzed the relationship of *CLE*s among three *Brassica* species and displayed the similarities and differences of *cis*-acting elements, interacted proteins, and expression patterns. With the aid of a natural population of *B. napus*, the genetic variations of *CLE* genes were uncovered, and several *CLE*s were recognized as candidate genes for important agronomic traits in rapeseed by regional association mapping analysis.

Polyploidization, an important force in the evolution of species (especially magnoliophyte), played an important role in plant adaptation o to new environments (Schemske, 1998; Zhang et al., 2020). Each Brassicaceae genome underwent a WGD event ~35 MYA ago (Bowers, 2003; Yuannian, 2011). Comparative genomic research showed that *Brassica* species experienced triploidy at the genomic level after they diverged from the *Arabidopsis* lineage approximately 20 MYA ago (Chalhoub et al., 2014; Liu et al., 2014). Therefore, it is apparent that the *Brassica* genome underwent paleopolyploidization (Bowers, 2003; Lysak et al., 2005). The differentiation of *B. rapa* and *B. oleracea* occurred approximately 4.6 MYA ago (Liu et al., 2014), and their natural hybridization formed the *B. napus* about 7,500 years ago (Chalhoub et al., 2014). Segmental duplication also leads to increased gene numbers (Flagel and Wendel, 2009). Most plants have undergone polyploidization events and thus retain a large number of duplicated chromosomal blocks at the genome level (Cannon et al., 2004). For example, in *B. napus*, segmental duplication/WGT is the primary force for *WOX* expansion (Li M. et al., 2019). In this study, nearly 80% of *CLE*s were distributed in syntenic blocks, suggesting that segmental duplication/WGT might contribute significantly to *CLE* expansion in *Brassica*. Tandem duplicated genes were defined as an array of at least two homologous genes within 50 kb (Cannon et al., 2004). However, we only found one pair of tandem duplicated genes in *B. rapa* and *B. oleracea*, suggesting that it is not a major factor for *CLE* expansion in *Brassica*.

Not all duplicated genes are retained in plants; gene loss always occurred due to the genomic sequence rearrangement after hybridization or chromosome doubling (Paterson et al., 2004; Ye et al., 2020). Each *AtCLE* should have three syntenic orthologs in *B. rapa* and *B. oleracea* based on the triploidy hypothesis. However, only *AtCLE5/6/45* had such a pattern, and the other *AtCLE*s had

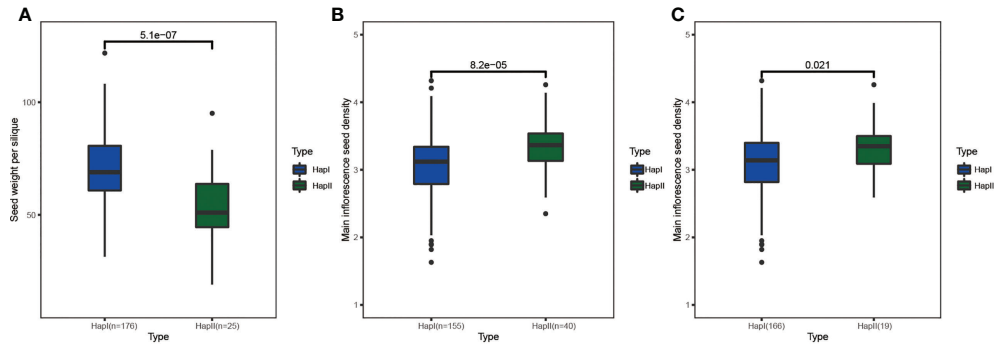


FIGURE 8
The haplotype analysis of *BnCLE*s for seed weight per silique and main inflorescence seed density. **(A)** *BnaC.CLV3.a*. **(B)** *BnaA.CLE20.b*. **(C)** *BnaC.CLE20.b*.

only one or two (less than three syntenic orthologs) in both species (Table S3). Hybridization of *B. rapa* and *B. oleracea* should theoretically result in six copies for each homologous *AtCLE* gene in *B. napus*, and the *CLE* number in *B. napus* is equal to the sum of the homologous genes in *B. rapa* and *B. oleracea*. In fact, only *AtCLE9* has six homologous genes in *B. napus*. Generally, most *CLE*s were lost during the formation of *B. napus*. Orthologs of nine *AtCLE*s were not found in three *Brassica* species. *B. napus* lost 38.02% of genes compared with *A. thaliana* at the genome-wide level (27,169 × 6 *vs.* 101,040), while nearly 63.54% of *CLE*s were lost in *B. napus*. The significantly higher average gene loss rate suggests that strong selection occurred for *CLE*s during evolution.

Theoretically, there were three possible means to lose syntenic *BnCLE*s. First, the loss of the *BrCLE*s or *BoCLE*s resulted in the loss of syntenic *BnCLE*s after *Brassica* genome triplication. *AtCLV3* has one syntenic ortholog (*BrCLV3* and *BoCLV3*) in *B. rapa* and *B. oleracea*, so it has two syntenic *BnCLV3*. Second, the loss of the *CLE*s occurred during the allopolyploidization process. *AtCLE2* has three syntenic orthologs in both *B. rapa* and *B. oleracea*, while it has four syntenic orthologs in *B. napus*. Finally, the two processes together caused the loss of syntenic *BnCLE*s. *AtCLE4* has one syntenic ortholog in both *B. rapa* and *B. oleracea*, whereas it has no syntenic ortholog in *B. napus*. These lost *CLE*s were redundant, and genetic change through loss could potentially lead to adaptive diversity.

Statistical analysis showed that 22 out of 29 *CLE* genes (76%) were positioned on the assembled chromosomes in *B. rapa*, whereas 12 out of 25 (48%) maintained their relative position in *B. oleracea* during the formation of *B. napus*. There are two possible reasons for this finding. One possibility is that the C_n subgenome has more abundant transposable elements (TEs) than the A_n subgenome (Chalhoub et al., 2014). The presence of TEs in the genome can cause the rearrangement of chromosomal sequences, which affects the genomic structure, including deletion, inversion, and translocation. Second, the C_n subgenome underwent more active homologous exchanges than the A_n subgenome during polyploidization (Chalhoub et al., 2014).

Only a few CLE peptides are studied in *Brassica*; however, the functions of most AtCLE peptides are characterized. Therefore, *Brassica* CLE functions can be predicted through sequence similarity and phylogenetic analysis. Seven distinct CLE subfamilies were defined using *Arabidopsis* and three *Brassica* species. For example, all *CLV3* genes are clustered into subfamily III, while *Brclv3* and *Bnclv3* cause enlarged meristems and lead to extra organs such as multilocular siliques, similar to the phenotype of *Atclv3* (Fletcher et al., 1999; Fan et al., 2014; Yang et al., 2018). This approach could elucidate the function of unknown CLE peptides. However, many genes contain more than one copy in polyploid species (Chalhoub et al., 2014; Li Z. et al., 2019). These duplicated genes are redundant, and variation in promoters could affect their gene expression (Davuluri et al., 2003; Klaas, 2009; Arsovski et al., 2015), For example, *BrCLE45* have three highly similar copies, but *cis*-elements have significant differences, so their expression patterns were different. Distinct expression patterns are a direct sign that duplicated genes may diverge in different directions, indicating the occurrence of non-functionalization, neo-functionalization, or sub-functionalization at the transcriptional level (Gallagher et al., 2016; Cheng et al., 2018). *BnCLE*s appear to be highly expressed in immature seeds, roots, and stems based on their expression levels in previous studies. Moreover, their interacting proteins are enriched in the development of these organs. Therefore, it is concluded that *CLE*s perform important functions in the biological process of these organs. However, the majority of *CLE*s hardly expressed in the tissues analyzed in this study. Therefore, they were either expressed in other developmental stages or conditions, or they were suppressed after duplication. Promoters that regulate the *CLE*s expression may also play a role. Abundant *cis*-elements involved in hormone response, development, and stress were identified, and their existence made the regulation more flexible, which helps plants adapt to complex environments.

Genetic variations in CLEs were investigated in the associated *B. napus* population; the smaller number of SNPs in CLEs indicated that this gene family sequence was very conserved. SNP density was

higher in the A subgenome compared with the C subgenome, which is consistent with other published gene families (Zhu et al., 2020; Wahid et al., 2022; Xie et al., 2022). Twelve *CLE*s were found to be significantly associated with important agronomic traits using regional association mapping analysis. Among them, the variations in the upstream region of *BnaC.CLV3.a* (the orthologs of the famous gene *CLV3*) were significantly associated with seed weight per silique. The sequence of variations was predicted as the binding sites of transcription factor SPL3, which can directly bind to the promoters of many genes to affect flower development (Jung et al., 2016). In *Brassica*, *BrCLV3* and *BnCLV3* control multilocular silique traits and increase seed production (Fan et al., 2014; Yang et al., 2018). Although no variation was detected in the genetic region of *BnaC.CLV3.a* in this natural population, the upstream variations that occurred in the binding sites of transcription factor SPL3 may influence the binding ability, which would possibly regulate the expression of *BnaC.CLV3.a* and affect the yield trait. In summary, the association mapping analysis conducted in this study could provide a better way to explore the significance of *CLE*s in phenotypic variation and offer candidate genes for further genetic improvement breeding in *B. napus*.

Conclusion

A total of 29, 32, and 70 *CLE*s were identified in *B. rapa*, *B. oleracea*, and *B. napus*, respectively, and divided into seven subfamilies in the phylogenetic tree. The conservation of the CLE domain suggested that the *CLE* family is relatively conserved in its biological function. WGT and segmental duplication were the major contributors to the expansion of *CLE*s in *Brassica* species. Transcriptome and qPCR analyses indicated that *BnCLE*s were highly expressed in immature seeds, roots, and stems. Some *CLE* pairs exhibited different expression patterns in the same tissue, indicating duplicated *CLE* differentiation. Moreover, genetic variations and regional association mapping analysis indicated that 12 *CLE* genes were potential genes for regulating important agronomic traits. In summary, this study was helpful in understanding the molecular evolution and biological function of *CLE*s in rapeseed and its diploid progenitors, which would aid future genetic improvement for high-yield breeding in *B. napus*.

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

YX, MX, and CZ designed the research, analyzed the data, and wrote the manuscript. MS analyzed the data. YX and CT reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by National Natural Science Foundation of China (32070217), Precursor projects of Guizhou province for biological breeding supporting by science and technology in 2022 (Fine identification and evaluation of crop germplasm resources), Subsidy project from NSFC of Guizhou Academy of Agricultural Sciences (No. [2021] 50), Science and technology project of Shandong Education Department to MS (Grant no. J15LE02), China Postdoctoral Science Foundation funded project to MS (Grant no. 2018M632646), the Young Top-notch Talent Cultivation Program of Hubei Province for CT, and the National Natural Science Foundation of China (31770250).

Acknowledgments

We thank Zhixian Qiao of the Analysis and Testing Center at IHB for the technical support in RNA-seq analysis.

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.2022.998082/full#supplementary-material

References

Allender, C., and King, G. (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

Arsovski, A. A., Pradinuk, J., Guo, X. Q., Wang, S., and Adams, K. L. (2015). Evolution of cis-regulatory elements and regulatory networks in duplicated genes of *Arabidopsis*. *Plant Physiol.* 169 (4), 2982–2991. doi: 10.1104/pp.15.00717

Beitz, E. (2000). T(E)Xshade: shading and labeling of multiple sequence alignments using (LTEX)-T-A 2(epsilon). *Bioinformatics* 16 (2), 135–139. doi: 10.1093/bioinformatics/16.2.135

Bidadi, H., Matsuoka, K., Sage-Ono, K., Fukushima, J., Pitaksaringkarn, W., Asahina, M., et al. (2014). CLE6 expression recovers gibberellin deficiency to promote shoot growth in *Arabidopsis*. *Plant J.* 78 (2), 241–252. doi: 10.1111/tpj.12475

Bolger, A. M., Marc, L., and Bjoern, U. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 15, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bonello, J., Opsahl-Ferstad, H., Perez, P., Dumas, C., and Rogowsky, P. (2000). Esr genes show different levels of expression in the same region of maize endosperm. *Gene* 246, 219–227. doi: 10.1016/s0378-1119(00)00088-3

Bowers, J. E. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 6930, 433-438. doi: 10.1038/nature01521

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4 (1), 10. doi: 10.1186/1471-2229-4-10

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

Chen, C., Chen, H., Zhang, Y., Thomas, H., Frank, M., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4 (5), 258–268. doi: 10.1038/s41477-018-0136-7

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6 (2), 80–92. doi: 10.4161/fly.19695

Czyzewicz, N., Shi, C. L., Vu, L. D., Van De Cotte, B., Hodgman, C., Butenko, M. A., et al. (2015). Modulation of *Arabidopsis* and monocot root architecture by CLAVATA3/EMBRYO SURROUNDING REGION 26 peptide. *J. Exp. Bot.* 66 (17), 5229–5243. doi: 10.1093/jxb/erv360

Davuluri, R. V., Sun, H., Palaniswamy, S. K., Matthews, N., Molina, C., Kurtz, M., et al. (2003). AGRIS: Arabidopsis gene regulatory information server, an information resource of *Arabidopsis* cis -regulatory elements and transcription factors. *BMC Bioinf.* 4, 25. doi: 10.1186/1471-2105-4-25

Dong, W., Wang, Y., and Takahashi, H. (2019). CLE-CLAVATA1 signaling pathway modulates lateral root development under sulfur deficiency. *Plants (Basel)* 8 (4), 103. doi: 10.3390/plants8040103

Fan, C., Wu, Y., Yang, Q., Yang, Y., Meng, Q., Zhang, K., et al. (2014). A novel single-nucleotide mutation in a CLAVATA3 gene homolog controls a multilocular silique trait in *Brassica rapa* l. *Mol. Plant* 7 (12), 1788–1792. doi: 10.1093/mp/ssu090

Fiers, M., Golemiec, E., van der Schors, R., van der Geest, L., Li, K., Stiekema, W., et al. (2006). The CLAVATA3/ESR motif of CLAVATA3 is functionally independent from the nonconserved flanking sequences. *Plant Physiol.* 141 (4), 1284–1292. doi: 10.1104/pp.106.080671

Fiers, M., Hause, G., Boutilier, K., Casamitjana-Martinez, E., Weijers, D., Offringa, R., et al. (2004). Mis-expression of the CLV3/ESR-like gene CLE19 in *Arabidopsis* leads to a consumption of root meristem. *Gene* 327 (1), 37–49. doi: 10.1016/j.gene.2003.11.014

Fiume, E. (2010). Expression analysis of the CLE signaling gene family in arabidopsis thaliana and functional characterization of CLE8 in seed development. *Electronic Thesis Dissertations*.

Fiume, E., and Fletcher, J. (2012). Regulation of *Arabidopsis* embryo and endosperm development by the polypeptide signaling molecule CLE8. *Plant Cell* 24 (3), 1000–1012. doi: 10.1105/tpc.111.094839

Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183 (3), 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Fletcher, J. C. (2020). Recent advances in arabidopsis CLE peptide signaling. *Trends Plant Sci.* 25, 1005-1016. doi: 10.1016/j.tplants.2020.04.014

Fletcher, J., Brand, U., Running, M., Simon, R., and Meyerowitz, E. (1999). Signaling of cell fate decisions by CLAVATA3 in *Arabidopsis* shoot meristems. *Science* 283 (5409), 1911–1914. doi: 10.1126/science.283.5409.1911

Gallagher, J., Grover, C., Hu, G., and Wendel, J. (2016). Insights into the ecology and evolution of polyploid plants through network analysis. *Mol. Ecol.* 25 (11), 2644–2660. doi: 10.1111/mec.13626

Gancheva, M. S., Dodueva, I. E., Lebedeva, M. A., Tvorogova, V. E., Tkachenko, A. A., and Lutova, L. A. (2016). Identification, expression, and functional analysis of CLE genes in radish (Raphanus sativus l.) storage root. *BMC Plant Biol.* 16 Suppl 1, 7. doi: 10.1186/s12870-015-0687-y

Grienenberger, E., and Fletcher, J. C. (2015). Polypeptide signaling molecules in plant development. *Curr. Opin. Plant Biol.* 23, 8–14. doi: 10.1016/j.pbi.2014.09.013

Han, S., Khan, M. H. U., Yang, Y., Zhu, K., Li, H., Zhu, M., et al. (2020). Identification and comprehensive analysis of the CLV3/ESR-related (CLE) gene family in brassica napus l. *Plant Biol. (Stuttg)* 22 (4), 709–721. doi: 10.1111/plb.13117

Han, H., Zhang, G., Wu, M., and Wang, G. (2016). Identification and characterization of the populus trichocarpa CLE family. *BMC Genomics* 17, 174. doi: 10.1186/s12864-016-2504-x

Hastwell, A., Gresshoff, P., and Ferguson, B. (2015). Genome-wide annotation and characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (Glycine max) and common bean (Phaseolus vulgaris), and their orthologues of arabidopsis thaliana. *J. Exp. Bot.* 66 (17), 5271–5287. doi: 10.1093/jxb/erv351

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* 44 (W1), W236–W241. doi: 10.1093/nar/gkw370

Hu, B., Jin, J., Guo, A., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinf. (Oxford England)* 31 (8), 1296–1297. doi: 10.1093/bioinformatics/btu817

Jung, J. H., Lee, H. J., Ryu, J. Y., and Park, C. M. (2016). SPL3/4/5 integrate developmental aging and photoperiodic signals into the FT-FD module in arabidopsis flowering. *Mol. Plant* 9 (12), 1647–1659. doi: 10.1016/j.molp.2016.10.014

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42 (4), 348–354. doi: 10.1038/ng.548

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Klaas, V. (2009). Unraveling transcriptional control in arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol.* 2, 535–536. doi: 10.1104/pp.109.136028

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19 (9), 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi: 10.1093/molbev/msw054

Larkin, M. (2007). Clustal W and clustal X v. 2.0. *Bioinformatics* 23 (21), 2947–2948. doi: 10.1093/bioinformatics/btm404

Letunic, I., Khedkar, S., and Bork, P. (2020). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937

Li, Z., Liu, D., Xia, Y., Li, Z., Niu, N., Ma, S., et al. (2019). Identification and functional analysis of the CLAVATA3/EMBRYO SURROUNDING REGION (CLE) gene family in wheat. *Int. J. Mol. Sci.* 20, 4317. doi: 10.3390/ijms20174319

Liu, L., Gallagher, J., Arevalo, E. D., Chen, R., Skopelitis, T., Wu, Q., et al. (2021). Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize CLE genes. *Nat. Plants* 7 (3), 287–294. doi: 10.1038/s41477-021-00858-5

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I., et al. (2014). The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5, 3930. doi: 10.1038/ncomms4930

Li, M., Wang, R., Liu, Z., Wu, X., and Wang, J. (2019). Genome-wide identification and analysis of the WUSCHEL-related homeobox (WOX) gene family in allotetraploid brassica napus reveals changes in WOX genes during polyploidization. *BMC Genomics* 20 (1), 317. doi: 10.1186/s12864-019-5684-3

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

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe brassiceae. *Genome Res.* 15 (4), 516–525. doi: 10.1101/gr.3531105

Magali, L. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Nekrutenko, A. (2002). The K A/K s ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. *Genome Res.* 12 (1), 198–202. doi: 10.1101/gr.200901

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. *Nat. Commun.* 4, 2191. doi: 10.1038/ncomms3191

Ostergaard, L., and King, G. J. (2008). Standardized gene nomenclature for the brassica genus. *Plant Methods* 4 (1), 10–10. doi: 10.1186/1746-4811-4-10

Paterson, A., Bowers, J., and Chapman, B. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U. S. A.* 101 (26), 9903–9908. doi: 10.1073/pnas.0307901101

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Rojo, E., Sharma, V., Kovaleva, V., Raikhel, N., and Fletcher, J. (2002). CLV3 is localized to the extracellular space, where it activates the arabidopsis CLAVATA stem cell signaling pathway. *Plant Cell* 14 (5), 969–977. doi: 10.1105/tpc.002196

Schemske, R. D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Systematics* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F. X., Jürgens, G., and Laux, T. (2000). The stem cell population of arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* 100 (6), 635–644. doi: 10.1016/S0092-8674(00)80700-X

Schranz, M. E., Lysak, M. A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the brassicaceae: building blocks of crucifer genomes. *Trends Plant ence* 11 (11), 535–542. doi: 10.1016/j.tplants.2006.09.002

Song, X., Guo, P., Ren, S., Xu, T., and Liu, C. (2013). Antagonistic peptide technology for functional dissection of CLV3/ESR genes in arabidopsis. *Plant Physiol.* 161 (3), 1076–1085. doi: 10.1104/pp.112.211029

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., et al. (2018). A small peptide modulates stomatal control *via* abscisic acid in long-distance signalling. *Nature* 556 (7700), 235–238. doi: 10.1038/s41586-018-0009-2

Tatsuhiko, K. (2006). A plant peptide encoded by CLV3 identified by *in situ* MALDI-TOF MS analysis. *Sci. (New York N.Y.)* 5788, 845–848. doi: 10.1126/science.1128439

Wahid, S., Xie, M., Sarfraz, S., Liu, J., Zhao, C., Bai, Z., et al. (2022). Genome-wide identification and analysis of Ariadne gene family reveal its genetic effects on agronomic traits of brassica napus. *Int. J. Mol. Sci.* 23, 6265. doi: 10.3390/ijms23116265

Wang, J., Replogle, A., Hussey, R., and Baum, T. (2011). Identification of potential host plant mimics of CLAVATA3ESR (CLE)-like peptides from the plant-parasitic nematode heterodera schachtii. *Mol. Plant Pathol.* 12 (2), 177–186. doi: 10.1111/j.1364-3703.2010.00660.x

Wang, X., Wu, J., Liang, J., Cheng, F., and Wang, X. (2015). Brassica database (BRAD) version 2.0: Integrating and mining brassicaceae species genomic resources. *Database J. Biol. Database Curation* 2015, bav093. doi: 10.1093/database/bav093

Xie, M., Zuo, R., Bai, Z., Yang, L., Zhao, C., Gao, F., et al. (2022). Genome-wide characterization of Serine/Arginine-rich gene family and its genetic effects on agronomic traits of brassica napus. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.829668

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, S., et al. (2018). Precise editing of CLAVATA genes in brassica napus l. regulates multilocular silique development. *Plant Biotechnol. J.* 16 (7), 1322–1335. doi: 10.1111/pbi.12872

Ye, C., Wu, D., Mao, L., Jia, L., Qiu, J., Lao, S., et al. (2020). The genomes of the allohexaploid echinochloa crus-galli and its progenitors provide insights into polyploidization-driven adaptation. *Mol. Plant* 13 (9), 1298–1310. doi: 10.1016/j.molp.2020.07.001

Yuannian, J. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 7345, 97-100. doi: 10.1038/nature09916

Zhang, L., Wu, S., Chang, X., Wang, X., Zhao, Y., Xia, Y., et al. (2020). The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant Cell Environ.* 43 (12), 2847–2856. doi: 10.1111/pce.13898

Zhang, Y., Yang, S., Song, Y., and Wang, J. (2014). Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. *BMC Genomics* 15 (1), 827. doi: 10.1186/1471-2164-15-827

Zhao, C., Safdar, L. B., Xie, M., Shi, M., Dong, Z., Yang, L., et al. (2021). Mutation of the PHYTOENE DESATURASE 3 gene causes yellowish-white petals in *Brassica napus*. *Crop J.* 9, 1124–1134. doi: 10.1016/j.cj.2020.10.012

Zhao, C., Xie, M., Liang, L., Yang, L., Han, H., Qin, X., et al. (2022). Genome-wide association analysis combined with quantitative trait loci mapping and dynamic transcriptome unveil the genetic control of seed oil content in brassica napus l. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.929197

Zhu, W., Guo, Y., Chen, Y., Wu, D., and Jiang, L. (2020). Genome-wide identification, phylogenetic and expression pattern analysis of GATA family genes in brassica napus. *BMC Plant Biol.* 20 (1), 543. doi: 10.1186/s12870-020-02752-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fpls.2022.1056206

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Fei Liu,
Henan University, China
Lunwen Qian,
Hunan Agricultural University, China
Yuanyuan Zhang,
Oil Crops Research Institute (CAAS), China

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

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 03 October 2022
ACCEPTED 26 October 2022
PUBLISHED 10 November 2022

CITATION
Zhang X, Li X, Li H, Wang Z, Xia R, Hu J, Wang P, Zhou X, Wan L, Hong D and Yang G (2022) Quantitative trait locus mapping and improved resistance to sclerotinia stem rot in a backbone parent of rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 13:1056206. doi: 10.3389/fpls.2022.1056206

Quantitative trait locus mapping and improved resistance to sclerotinia stem rot in a backbone parent of rapeseed (*Brassica napus* L.)

Xiaohui Zhang[1,2,3], Xiang Li[1], Huining Li[1], Zhuanrong Wang[4], Rui Xia[1], Jin Hu[2,3], Pengfei Wang[1], Xianming Zhou[2,3], Lili Wan[4], Dengfeng Hong[1]* and Guangsheng Yang[1,2,3]*

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China, [2]Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Sanya, China, [3]College of Tropical Crops, Hainan University, Haikou, China, [4]Institute of Crops, Wuhan Academy of Agricultural Sciences, Wuhan, China

There are three main challenges to improving sclerotinia stem rot (SSR) resistance in rapeseed (*Brassica napus* L.). First, breeding materials such as the backbone parents have not been extensively investigated, making the findings of previous studies difficult to directly implement. Second, SSR resistance and flowering time (FT) loci are typically linked; thus, use of these loci requires sacrifice of the rapeseed growth period. Third, the SSR resistance loci in susceptible materials are often neglected, thereby reducing the richness of resistant resources. This study was conducted to investigate the stem resistance, disease index, and FT of a doubled haploid population consisting of 151 lines constructed from the backbone parent 19514A and conventional rapeseed cultivar ZY50 within multiple environments. Quantitative trait locus (QTL) mapping revealed 13 stem resistance QTLs, 9 disease index QTLs, and 20 FT QTLs. QTL meta-analysis showed that *uqA04*, *uqC03.1*, and *uqC03.2* were repeatable SSR resistance QTLs derived from different parents but not affected by the FT. Based on these three QTLs, we proposed a strategy for improving the SSR resistance of 19514A and ZY50. This study improves the understanding of the resistance to rapeseed SSR and genetic basis of FT and demonstrates that SSR resistance QTLs can be mined from parents with a minimal resistance level difference, thereby supporting the application of backbone parents in related research and resistance improvement.

KEYWORDS

quantitative trait locus mapping, sclerotinia stem rot, flowering time, backbone parent, rapeseed

Introduction

Rapeseed (*Brassica napus* L.) is among the most important oil crops worldwide and its supply affects the stability of edible oil production. The yield of rapeseed is affected by several factors, such as yield potential, flowering time (FT), and yield stability. Poor yield stability is the greatest contributor to low yields. Sclerotinia stem rot (SSR) is a fatal disease in plants and is caused by the typical dead trophic pathogen *Sclerotinia sclerotiorum*. This disease is prevalent during rapeseed growth and can cause serious yield losses (Bolton et al., 2006). In China, SSR occurs in all rapeseed planting areas, particularly in areas where winter rapeseed is grown, resulting in related annual yield losses of 10–20%. In areas severely affected by SSR, yield loss can exceed 80% or lead to complete harvest loss (Xu et al., 2014; Yu et al., 2020; Ding et al., 2021). SSR can also reduce the oil content of rapeseed seeds and affect the quality of rapeseed oil, which may negatively impact human health (Pressete et al., 2019). Compared with chemical prevention and control, improving the SRR resistance and breeding resistant cultivars of rapeseed are more economical, efficient, and sustainable methods for mitigating damage caused by SSR (Ding et al., 2021).

The basic requirements for SSR resistance breeding are to establish efficient and accurate identification methods and search for resistant germplasm resources. Various methods for SSR resistance identification have been proposed and are widely used at the seedling, flowering termination, and mature stages; however, no germplasm resources with high resistance or complete immunity have been identified (Ding et al., 2021). Although a series of genes was confirmed *via* reverse genetics to be involved in regulating SSR resistance (Wang et al., 2014; Wang et al., 2019a; Jiang et al., 2020; Wang et al., 2020; Cao et al., 2022; Zuo et al., 2022) and progress has been made in understanding its regulatory network (Cao et al., 2016; Wang et al., 2019b; Hu et al., 2021; Xu et al., 2021; Zhang et al., 2022), it is difficult to apply these results in the short term because of the limited material specificity and transgenic restriction policies. Thus, the main strategy used to improve SSR resistance is constructing isolated populations and analyzing the genetic basis of SSR resistance of two inbred lines showing a large discrepancy in their SSR resistance levels, followed by exploration and utilization of QTLs related to SSR resistance. Backbone parents with a high combining ability often produce excellent hybrid cultivars (Fradgley et al., 2019; Ma et al., 2019; Chen et al., 2021). SSR resistance of the backbone parents and improvements in their SSR resistance have been studied at the breeding level, and the results can be rapidly applied in breeding. However, in this strategy, a large difference in the SSR resistance level is the main criterion for selecting parents, and backbone breeding parents that have undergone natural selection and strict artificial selection are rarely used as susceptible parents because their SSR resistance levels are not low enough. Using this strategy, numerous researchers have applied various methods for detecting SSR to discover a large number of SSR resistance-related QTLs in different growth stages of rapeseed (Zhao et al., 2006; Yin et al., 2009; Mei et al., 2013; Wu et al., 2013; Li et al., 2015; Wei et al., 2016; Wu et al., 2016a; Shao et al., 2022). However, the correlation between the results of different identification methods was low, as was the repeatability of QTLs mined in different studies; therefore, few or no QTLs have been applied to improve resistance to SSR (Ding et al., 2021). Nevertheless, these studies repeatedly confirmed that SSR resistance is not controlled by a single major locus but rather by multiple micro loci, thus revealing the genetic pattern of SSR resistance. Moreover, studies demonstrated that even relatively susceptible parents can provide a source of resistance (Wu et al., 2013). Thus, breeding materials can be directly used to construct populations for exploring resistance loci as a potential strategy to improve the SSR resistance of the backbone parents regardless of resistance differences in the parents.

SSR resistance in rapeseed is related to the stem strength (Shao et al., 2022), lignin content (Cao et al., 2022), glucosinolate content (Zhao and Meng, 2008), FT (Wei et al., 2014; Wu et al., 2019; Zhang et al., 2019), and other traits (Feng et al., 2021), and their interactions should be monitored during the improvement of SSR resistance. A study of the FT and SSR resistance of 521 rapeseed inbred lines revealed a significant negative correlation between the two traits, demonstrating that they had the closest relationship. Inbred lines with early FT were more susceptible to SSR, whereas those with late FTs were more resistant to SSR (Zhang et al., 2019). Furthermore, several studies showed that QTLs for FT colocalized with those for SSR resistance, confirming the genetic linkage between the FT and SSR resistance loci (Wei et al., 2014; Wu et al., 2019; Zhang et al., 2019). The pleiotropic effect of a single gene is one reason for the linkage between the FT and SSR resistance loci in *Arabidopsis thaliana* (Kidd et al., 2009; Li et al., 2012; Lyons et al., 2013; Singh et al., 2013; Lai et al., 2014), but this has not been clearly demonstrated in rapeseed. FT has a comprehensive impact on rapeseed cultivars, and differences in the FT directly affect the adaptability of rapeseed cultivars to different planting areas as well as yield (Raman et al., 2019; Kaur et al., 2021). Rapeseed improvement requires shorter growth times and earlier flowering periods. Therefore, improving the SSR resistance of rapeseed should not be at the expense of FT. When analyzing the QTL for SSR resistance and using the results to guide improvements in breeding parents, the QTL for FT should be avoided.

In this study, the temperature-sensitive pol cytoplasmic male sterile line 19514A (a backbone parent) and conventional cultivar ZY50, along with their constructed doubled haploid (DH) population containing 151 lines, were used to examine the resistance to SSR using a stem resistance (SR) assay at the flowering termination stage and disease index (DI) at the mature stage for two consecutive years. The SR QTLs and DI QTLs were explored based on a high-density genetic linkage map. Additionally, the phenotypes of the FT of the parents and

their DH populations were investigated in multiple environments, and the genetic basis of the differences in the FT of the DH population was analyzed. Through colocalization analysis of the QTL for SR, DI, and FT, a SSR resistance improvement scheme for 19514A and ZY50 was proposed. This study improves the understanding of the genetic basis of SSR resistance and FT in rapeseed and provides useful information for improving SSR resistance in backbone parents and SSR resistance of 19514A and ZY50.

Materials and methods

Plant material and growth conditions

19514A (named G120) is a temperature-sensitive pol cytoplasmic male sterile line. This semi-winter backbone sterile line has been formulated with many hybrid cultivars that are promoted in the market but shows a low SSR level. ZY50 (named as 9172) is a conventional semi-winter rapeseed cultivar with excellent performance. The DH population constructed with ZY50 and 19514A contains 151 lines, for which a high-density genetic linkage map containing 910 single-nucleotide polymorphism markers and 187 simple sequence repeats markers was previously constructed (Liu et al., 2020). All plant materials were sown in the field of Wuhan (WH), Jingzhou (JZ), or Zhangye (ZY) during normal growing seasons. Each row consisted of 10–12 plants, with distances of 20 cm between individuals and 25 cm between rows. Conventional field management was conducted according to local planting practices.

Stem inoculation assay

The parents and their DH population were planted at the experimental base of Huazhong Agricultural University in WH for two consecutive years from 2015 to 2016; three rows of each material were planted. Inoculation was performed as described by Wang et al. (2018) and Wu et al. (2013). The *S. sclerotiorum* isolate SS-1 was maintained and cultured on potato dextrose agar (PDA, 25% potato, 2.5% dextrose and 1.5% agar, pH 5.8). The isolate was cultured more than two cycles prior to inoculation at 23°C in darkness. At the flowering termination stage of rapeseed, agar discs (8 mm in diameter) were excised from the edges of growing fungal colonies and up-ended into the lids of 1.5- or 2.0-mL centrifuge tubes. These tubes were affixed with plastic wrap onto rapeseed stems at 30 cm from the ground. Disease severity was assessed twice by measuring the lesion length per pathogen infection spot at 7 and 14 days post-infection (dpi) (17 dpi in 2015). For each DH line, 10 plants showing a consistent growth status were selected for inoculation, and the single plant in the middle row was preferentially selected to reduce the impact of marginal effects. In the collected data, the average value of the six-middle data of each family was used as the phenotype value.

Natural infection experiment

The parents and their DH population were planted in JZ for two consecutive years from 2015 to 2016. The experiments were performed in a randomized block design with two replicates, and each line was planted in three rows to ensure that there were approximately 30 plants. As described by Wang et al. (2018), five agar discs (8 mm in diameter) containing active mycelia were added to 500 ml of liquid potato medium (25% potato and 2.5% dextrose, pH 5.8) and incubated at 200 rpm for 3 days at 23°C in the dark. The mycelia were then fully interrupted and diluted to 10 L with water, and then the 5% suspension of *S. sclerotiorum* hyphae was sprayed onto the rapeseed plants at the full-flowering stage to increase disease-causing stress. Before harvest, the incidence level of each plant was evaluated as described by Zhou et al. (1993). The DI of each plot was calculated as $DI = 100\ \Sigma(i \times n_i)/(N \times k)$, where i is the disease severity score from 0 to 4, n_i is total number of plants in each score, N is total number of plants evaluated in each plot, and k is the highest score (here, k = 4).

FT evaluation

The parents and their DH population were planted in ZY for two consecutive years from 2015 to 2016; JZ in 2016; and WH in 2015, 2016 and 2018. The experiments were conducted in a randomized block design with two replicates, and each replicate was planted in three rows. The phenotypic values of individual plants in each line from the sowing date to 50% flowering were recorded as the flowering date.

Data statistics and analysis

Data collection and preliminary analysis were performed using Microsoft Excel 2016. One-way analysis of variance, correlation analysis, and graphical presentation of data were performed using GraphPad Prism 8 software (GraphPad, Inc., La Jolla, CA, USA). The phenotypic data generated in this study are shown in Supplementary Data 1.

QTL mapping and meta-analysis

QTL analysis was performed *via* composite interval mapping using WinQTL cartographer 2.5 software (Zeng, 1994). The walk speed was set to 1 cM. The limit of detection (LOD) threshold

for each trait was determined using permutation testing with 1000 repetitions. A QTL was declared when the LOD score was greater than the threshold value; LOD scores corresponding to P < 0.05 were used to identify significant QTLs. QTLs repeatedly detected in different environments and different trait were integrated into consensus QTLs through meta-analysis using BioMercator 2.1 software (Arcade et al., 2004).

Results

Phenotypic identification and QTL mapping for SR

To identify the SR of the parents and their DH population, we performed stem inoculation experiments at maturity for two consecutive years. We recorded the phenotype data at 7 days after inoculation (15WHSR-7D and 16WHSR-7D) and several days after inoculation (15WHSR-17D and 16WHSR-14D). The difference between the two sets of data were used as the third group of phenotype data (15WHSR-C and 16WHSR-C). The results showed that the SR of 19514A and ZY50 was not significantly different at all four phenotype collections over two years; however, stable and significant differences in SR were observed between specific DH lines (Figures 1A-C). In the DH population, SR showed a continuous distribution in all six datasets, with the performance of the parents in the middle of that of the DH population (Figures 1D–I). In correlation analysis, the three sets of data in the same year showed significant positive correlations. Across different years, except for the 15WHSR-7D and 16WHSR-7D groups, the other two groups of data also showed significant positive correlations (Supplementary Table 1). This result confirms the reliability of the experimental data and suggests that multiple loci regulate SR in the DH population, and that both parents can provide resistance sources.

To mine the loci controlling SR in the DH population, we combined the existing high-density genetic linkage map and these six datasets for QTL mapping. We detected 13 QTLs with LOD values of 2.65–4.84 and phenotypic variation of 4.3–11.2%. The additive effects of different QTLs were from different

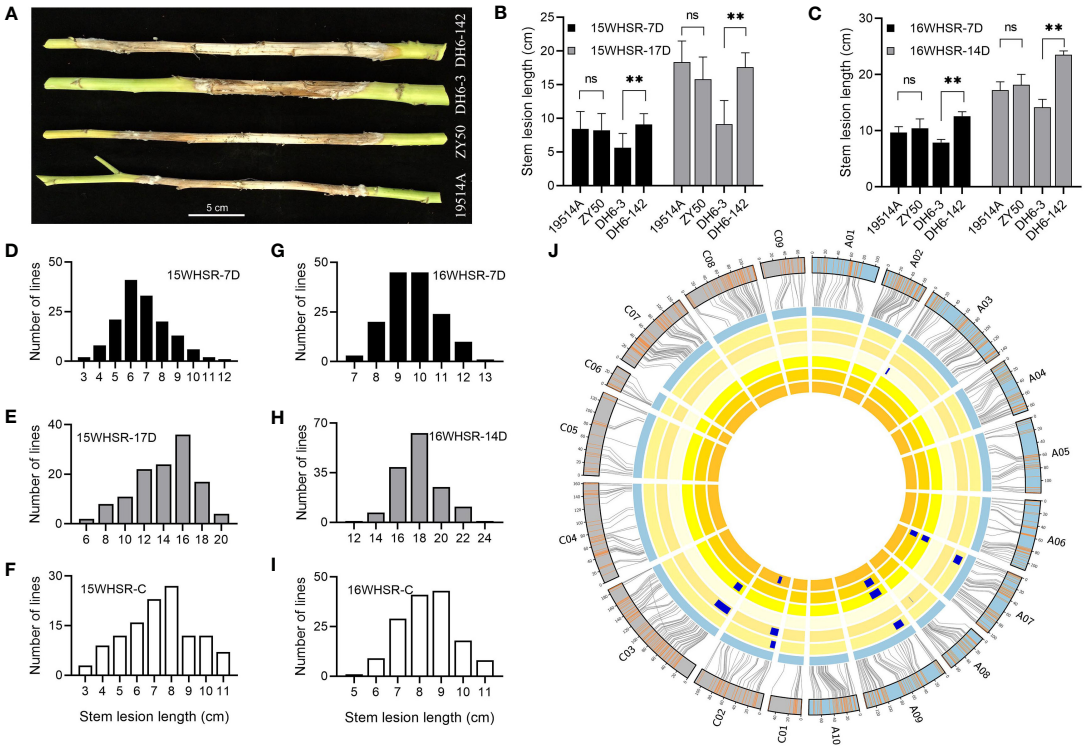


FIGURE 1
Phenotype and quantitative trait loci (QTL) mapping of stem resistance assay. **(A)** Disease lesions on the stems of the two parents and two double haploid (DH) lines at 14 days post-infection (dpi). Bar = 5 cm. **(B, C)** Stem lesion lengths of the two parents and two DH lines in 2015 and 2016. "-7D" indicates the data from the first measurement at 7 dpi and "-14D" and "-17D" indicate the data from the second measurement. Data are shown as the mean ± SD; ns indicates no significant difference, **P < 0.01 (one-way analysis of variance). **(D-I)** Distribution of stem lesion lengths of the DH population in 2015 **(D–F)** and 2016 **(G–I)**. "-C" indicates the difference between the two measurements. **(J)** Genetic linkage map and locations of QTLs for stem resistance (SR). From inside to outside, the six cycles represent 15WHSR-7D, 15WHSR-17D, 15WHSR-C, 16WHSR-7D, 16WHSR-14D, and 16WHSR-C. The two outermost cycles show a comparison of the linkage and physical maps of *Brassica napus*.

parents, indicating that both parents provide a source of resistance (Supplementary Table 2). The identified QTLs were distributed on the A03, A07, A08, A09, C02, and C03 linkage groups, and the confidence intervals of QTLs in different datasets overlapped (Figure 1J). Among these QTLs, *qSRA07-1* and *qSRA09-1* were detected repeatedly in 2015, and *qSRC02-2* was detected repeatedly in 2016, showing that different data collection methods can identify stable QTLs. *qSRA09-1* and *qSRC03-1* were detected across different years and considered as stable QTLs, and thus should be further evaluated.

Phenotypic identification and QTL mapping for DI

To measure the resistance of the parents and their DH population to SSR using the DI, we investigated the phenotypes of four replicates over two years. As with SR, DI did not significantly differ between the two parents, although there were stable significant differences between specific DH lines (Figures 2A, B). The DI of the DH population showed a large range and an approximately normal distribution, indicating that multiple loci in the DH population simultaneously regulate the DI (Figures 2C–F). Correlation analysis revealed significant positive correlations among the four replicates (Supplementary Table 3), demonstrating the stability of this method for identifying disease resistance and the reliability of the data.

A total of nine QTLs were detected in the four replicates, with LOD values ranging from 2.54 to 5.82, explaining 5.3% to 12.2% of the phenotypic variation. The additive effects of these QTLs ranged from 3.61 to 6.42, with resistance derived from different parents (Supplementary Table 4). According to the confidence intervals, six QTLs were integrated into three consensus QTLs on chromosomes A04, C02, and C03 (Figure 2G). These three QTLs showed similar phenotypic variation and additive effects, and thus may jointly regulate SSR resistance at the same level.

QTL mapping for FT

To mine the QTL regulating FT in the DH population, we investigated 10 replicates of FT in five environments. Correlation analysis showed significant positive correlations among the 10 replicates of FT data (Supplementary Table 5). In most environments, the FT of the DH population was distributed continuously over 20 days. In WH in 2015, the FT lasted for 30

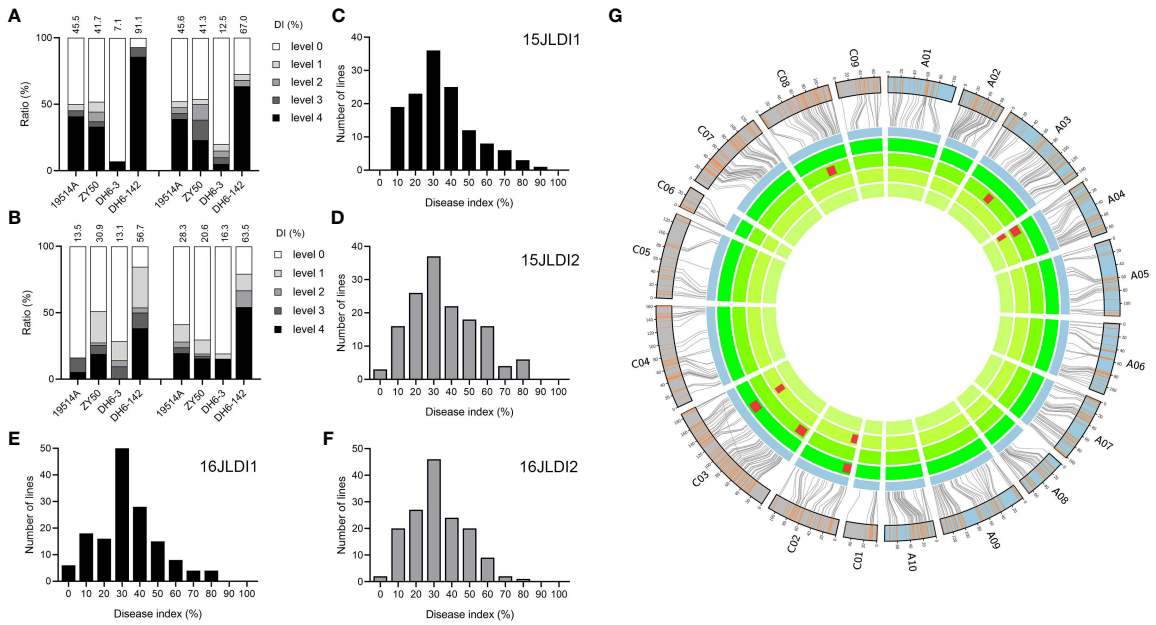


FIGURE 2
Phenotype and quantitative trait loci (QTL) mapping of natural infection experiment. **(A)** Disease indices (DIs) of the two parents and two double haploid (DH) lines in two replicates in 2015. The histograms represent the ratio of plants with different disease levels. **(B)** DIs of the two parental lines and two DH lines of two replicates in 2016. **(C–F)** Distribution of DIs of the DH population in 2015 **(C, D)** and 2016 **(E, F)**. **(G)** Genetic linkage map and locations of QTLs for DI. From inside to outside, the four cycles represent 15JZDI1, 15JZDI2, 16JZDI1, and 16JZDI2. The two outmost cycles show a comparison of the linkage and physical maps of *Brassica napus*.

days. None of the replicates showed an obvious Mendelian distribution, and most replicates did not conform to the standard normal distribution. These results indicate that the FT of the DH population is controlled by a major locus and multiple minor loci (Supplementary Figure 1).

The results of QTL mapping showed that 20 QTLs for FT were detected in 8 of 10 replicates, among which 3 and 17 QTLs were detected in the spring and winter rapeseed growth region, respectively (Figure 3). These QTLs were distributed on chromosomes A02, A04, A07, A09, A10, C02, and C03, with *qFTC02-1* detected in a total of six repeats in the three locations with LOD values of 2.53–3.04, 26.03–27.07, and 15.39, phenotypic contribution rates of 6.2–8.1%, 43.8–48.2%, and 27.6%, and additive effects of 1.65d–1.85d, 3.48d–5.27d, and 2.44d. Another QTL, *qFTA07-1*, was detected in five replicates in WH and JZ, among which four replicates except for 15WHFT1 showed LOD values of 2.77–4.38, phenotypic contributions of 3.3–4.9%, and additive effects of 0.93d–1.32d. In addition, the QTL *qFTA9-1* was detected in two replicates in WH (Supplementary Table 6). The above results confirm that the FT differed within the DH population and was simultaneously regulated by multiple loci. In WH and JZ, *qFTC02-1* was a stable major locus, whereas *qFTA07-1* and *qFTA09-1* were stable minor loci.

QTL meta-analysis for SR, DI, and FT

To analyze whether the previously detected QTLs for SSR resistance are affected by the FT, we performed a meta-analysis of all QTLs for the above three traits. The results showed that seven QTLs were repeatedly detected, among which *uqC02* was stably detected in all three traits and *uqA07* was detected in SR and FT. These results indicate that the two QTLs affect FT to influence the SSR resistance of rapeseed. *uqA09.1* and *uqA09.2* were repeatedly detected in SR and FT, respectively. Although they were identified as two different QTLs, their confidence intervals showed a small overlap; thus, we considered this QTL to play a role in SSR resistance while also affecting the FT. The remaining three QTLs, *uqA04*, *uqC03.1*, and *uqC03.2*, were stably detected in SR or DI and were not affected by the FT, although *uqA04* may be considered as a minor FT QTL in one replicate (Table 1). The above results indicate that the SSR resistance of this DH population was partly regulated

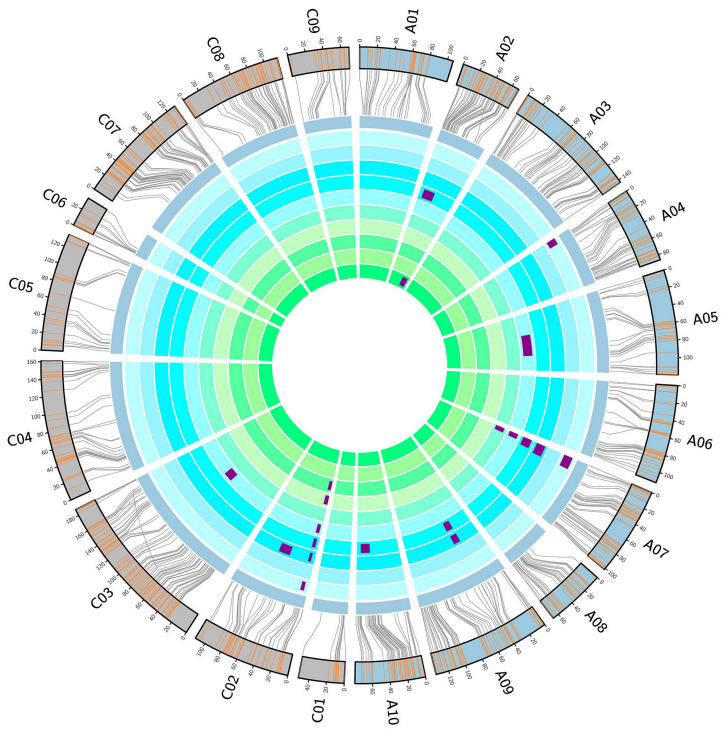


FIGURE 3
Genetic linkage map and locations of quantitative trait loci (QTL) for flowering time (FT). From inside to outside, the ten cycles represent 15ZYFT1, 15ZYFT2, 16ZYFT1, 16ZYFT2, 15WHFT1, 15WHFT2, 16WHFT1, 16WHFT2, 16JZFT1, and 16JZFT2. The two outmost cycles show a comparison of the linkage and physical maps of *Brassica napus*.

10.3389/fpls.2022.1056206

TABLE 1 Unique quantitative trait loci (QTL) information in this study.

Name	Chr	Position (cM)	CI (cM)	SR	DI	FT	References
uqA04	A04	4.19	0.47-7.91		15JZDI2/ 16JZDI1	16JZFT2	
uqA07	A07	2.65	0-5.54	15WH-17D/15WH-C		15WHFT1/ 15WHFT2/ 16WHFT1/ 16WHFT2/ 16JZFT2	
uqA09.1	A09	39.95	25.61-39.47	15WH-17D/15WH-C/ 16WH-C			Wu et al., 2013; Wei et al., 2014; Zhang et al., 2019
uqA09.2	A09	42.87	38.52-47.21			15WHFT1/ 15WHFT2	
uqC02	C02	8.64	7.35-9.92	16WH-14D/16WH-C	15JZDI2/ 16JZDI2	16ZYFT1/16ZYFT2/ 16WHFT1/ 16WHFT2/ 16JZFT2	Wei et al., 2014; Wu et al., 2019; Zhang et al., 2019
uqC03.1	C03	22.39	14.76-30.01	15WH-C/16WH-14D	16JZDI1		Wu et al., 2019
uqC03.2	C03	130.79	124.26-137.33		15JZDI2/ 16JZDI2		

Chr, chromosome; SR, stem resistance; DI, disease index; FT, flowering time.

by the FT. However, we identified several SSR resistance QTLs that were not related to the FT, which should be further evaluated.

Feasible strategies for improving SSR resistance of parents

To explore the application value of the improved SSR resistance of *uqC02*, *uqA04*, *uqC03.1*, and *uqC03.2*, we conducted haplotype analysis of DI in the DH population. Four QTLs divided the 151 DH lines into 16 haplotypes (Figure 4A). The obvious differences in the FTs of haplotype 1–8 and haplotype 9–16 suggest that the regulation of SSR resistance of the other three QTLs should be considered based on *uqC02*. When the genotype of *uqC02* was consistent with that of 19514A, haplotype 13 showed the highest DI in all four replicates, whereas its complementary haplotype 12 exhibited a ubiquitously low DI. The difference between the two haplotypes was very significant in three replicates. Additionally, the DI of haplotype 16, which was consistent with that of 19514A, was always between those of haplotype 13 and haplotype 12 (Figure 4B), possibly because of the source of additive effects. Interestingly, when the genotype of *uqC02* was consistent with that of ZY50, the above patterns were not significant (Figure 4C). We further compared haplotype 1 and haplotype 16 with the other haplotypes. The results showed that in all four replicates, the DI of haplotype 4 was lower than that of haplotype 1 with a difference of 7.8–12%, and the DI of haplotype 12 was smaller than that of haplotype 16, with a difference of 4.8–18.7% (Figure 4D). The above results confirm that *uqA04*, *uqC03.1*, and *uqC03.2* regulate SSR resistance, and that introduction of *uqC03.1* and *uqC03.2* from 19514A into ZY50 and *uqA04* from ZY50 into 19514A can enhance SSR resistance.

Discussion

With the rapid development of genotype analysis technology, phenotypes have become key factors in exploring valuable genetic loci. Using traditional methods, it was difficult to accurately identify the phenotype of SSR resistance in rapeseed, and the identifiable phenotype in the seedling stage did not always accurately reflect the final resistance level. Additionally, the phenotypes of plants cultivated in a greenhouse frequently do not reflect the phenotypes of those in the field, whose development is typically unstable because of the influence of environmental factors (Ding et al., 2021).

Most methods for stem inoculation used in previous studies involved measuring the lesion length within 7 days after inoculation (Zhao and Meng, 2003; Wei et al., 2014; Wei et al., 2016; Wu et al., 2016b; Wang et al., 2018; Wu et al., 2019; Zhang et al., 2019; Ding et al., 2020). In this study, we measured the lesion length a second time at 14 days after inoculation. Compared with the data at 7 dpi, the phenotypic values of the DH population at 14 dpi showed a wider range, the data showed better reproducibility, and more QTL for SSR resistance were identified. In addition, the increase in lesion length from 7 to 14 days was very reproducible after two years, better reflecting the

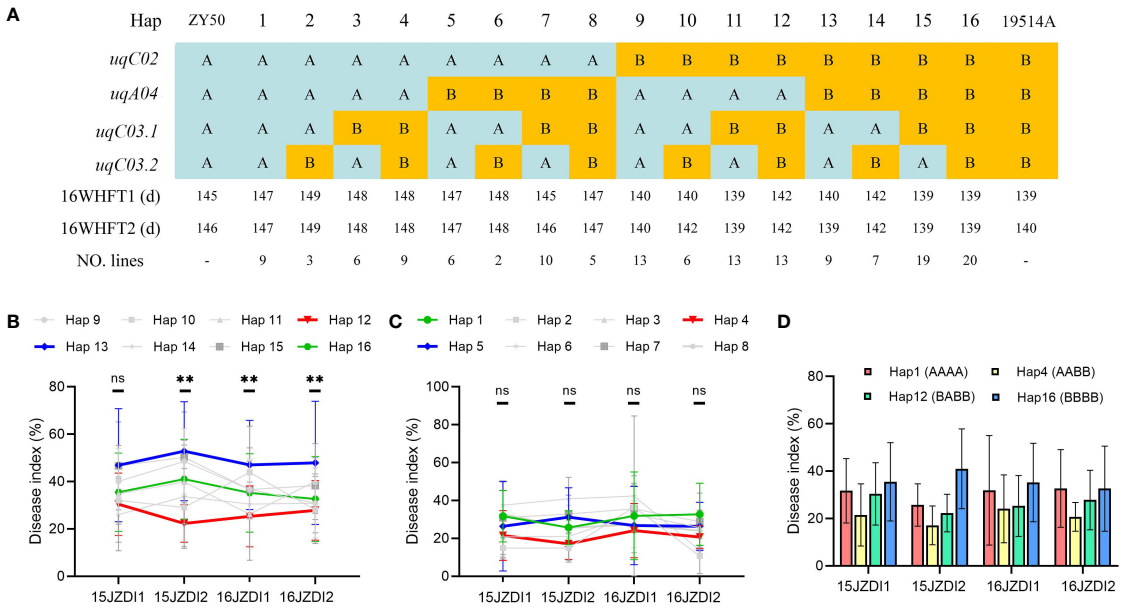


FIGURE 4
Haplotype analysis in the double haploid (DH) population. **(A)** DH population was divided into 16 haplotypes based on the genotypes of the four quantitative trait loci (QTLs). The number of lines per haplotype and average flowering time (FT) are shown. **(B)** Mean value of disease index **(DI)** and order of haplotypes 9–16 in four replicates. Data are shown as the mean ± SD. Differences between haplotype 12 and haplotype 13 were analyzed for significance; ns indicates no significant difference, **P < 0.01 (one-way analysis of variance). **(C)** Mean value of DI and order of haplotypes 1–8 in four replicates, with the data shown as the mean ± SD. Differences between haplotype 4 and haplotype 5 were analyzed for significance; ns indicates no significant difference (one-way analysis of variance). **(D)** DI values of haplotype 4 were less than those of haplotype 1 in all four replicates; the DI values of haplotype 12 were less than those of haplotype 16 in all four replicates. Data are shown as the mean ± SD.

resistance level of materials to SSR after successful infection. These results suggest that appropriately prolonging the inoculation time can reflect the differences in the materials' resistance levels, and that measuring the lesion length after 14 days of inoculation or the increase in lesion length from 7 to 14 days as the phenotypic value may more accurately reveal their resistance levels. Similar systematic problems were observed in many recent studies, and the infection times were appropriately extended (Zhang et al., 2019; Gupta et al., 2020).

Another method for identifying disease resistance used in this study is natural infection experiments. To increase pathogenic pressure, a 5% suspension of *S. sclerotiorum* hyphae was sprayed onto rapeseed plants at the full-flowering stage. The phenotypic data of four replicates in two years were significantly positively correlated, and QTL related to SSR resistance were repeatedly detected in multiple environments. These results demonstrate that the natural infection experiment is a relatively stable method for detecting SSR resistance; this method was also recently applied in numerous studies (Zhang et al., 2019).

Using the methods described above, we identified 13 SR QTLs, 9 DI QTLs, and 20 FT QTLs. Among them, *uqC02* was detected in all three traits, and is considered as a locus that simultaneously regulates SSR resistance and FT, which is consistent with the results of many previous reports (Wei et al., 2014; Wu et al., 2019; Zhang et al., 2019). Interestingly, *uqC02* was a major QTL that explained more than 40% of the phenotypic variation in the FT in WH and 27.6% of that in the FT in JL, showing a significantly stronger effect than those of other loci. However, *uqC02* only explained approximately 10% of the phenotypic variation in SSR resistance, which is comparable to the effect of other loci, and in this instance cannot be considered as a major QTL. This result suggests that the SSR resistance phenotype in this DH population is not completely controlled by the FT, and the application of other SSR QTL may improve SSR resistance without altering the FT. Notably, the gene regulating the FT in *uqC02* may have been cloned, and *BnaFLC.C2* is considered to have caused the variation in FT (Chen et al., 2018). Based on the close relationship between the FT and SSR resistance, *BnaFLC.C2* likely also regulates SSR resistance in *uqC02*. Although *uqC02* has some application value in improving the rapeseed growth period (Fang et al., 2021), its application value in SSR resistance improvement should be further investigated because of the link between the FT and SSR resistance.

uqA09.1 and *uqA09.2* were predicted to regulate SR and FT, respectively, but the confidence interval of the two loci showed

some overlap, and a single locus may control SSR resistance and FT simultaneously. Previous studies also demonstrated that these loci are responsible for SSR resistance (Wu et al., 2013; Wei et al., 2014; Zhang et al., 2019); however, because the interval is also an area of concentrated rapeseed yield genes (Liu et al., 2015; Shi et al., 2019), a more specific and elaborate design is required before this segment can be used to improve SSR resistance.

Regulation of the FT by *uqA07* was stably detected in the winter rapeseed environment but not in the spring rapeseed environment, indicating that *uqA07* is an environment-specific QTL for FT. As *uqA07* was involved in regulating SR but did not affect NI, it was not further analyzed nor applied in this study. However, *uqA07* may play an important role in finely regulating the FT in rapeseed, and *BnaFT.A07* may be the functional gene of *uqA07*.

Several other SSR resistance QTLs were also identified in this study, among which the resistance source of *uqA04* was ZY50 and that of *uqC03.1* and *uqC03.2* was 19514A. The phenotypic contribution and additive effect of the three QTLs were similar, and thus they were considered to have the same resistance levels. Previous studies reported *uqC03.1* as a SSR resistance QTL (Wu et al., 2019), whereas *uqA04* and *uqC03.2* were previously unreported. In this study, we confirmed that materials with significant differences in SSR resistance can be obtained by using a reasonable combination of these three loci. According to the source of resistance, application of *uqA04* may effectively improve the SSR resistance of the backbone parent 19514A. Our results confirm that low SSR resistance materials can also provide SSR resistance sources; we also demonstrated that the polymerization of SSR resistance loci in low SSR resistance materials is another effective strategy for improving rapeseed SSR resistance. These results can be applied when selecting SSR resistance cultivars in rapeseed. In addition, our results provide a foundation for further studies of major diseases in other crops and a reference for evaluating crop diseases.

Data availability statement

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

Author contributions

XHZ conducted most experiments and wrote the original draft. XL, HL, and RX participated in phenotypic data collection. PW and XMZ participated in data analysis. ZW, LW, and JH designed the experiments and were involved in reviewing and editing the manuscript. DH and GY supervised the project. All authors read and contributed to the revision of manuscript.

Funding

This research was supported by the Program for Modern Agricultural Industrial Technology System (CARS-12), Open Fund of the National Key Laboratory of Crop Genetic Improvement (ZK201909), and scientific research start funds of Hainan University.

Acknowledgments

We would like to thank Editage (www.editage.cn) for English language editing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1056206/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of the flowering time (FT) of double haploid (DH) populations in ten replicates.

References

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 (14), 2324–2326. doi: 10.1093/bioinformatics/bth230

Bolton, M. D., Thomma, B. P., and Nelson, B. D. (2006). *Sclerotinia sclerotiorum* (Lib.) de bary: biology and molecular traits of a cosmopolitan pathogen. *Mol. Plant Pathol.* 7 (1), 1–16. doi: 10.1111/j.1364-3703.2005.00316.x

Cao, J. Y., Xu, Y. P., and Cai, X. Z. (2016). TMT-based quantitative proteomics analyses reveal novel defense mechanisms of *Brassica napus* against the devastating necrotrophic pathogen *Sclerotinia sclerotiorum*. *J. Proteomics* 143, 265–277. doi: 10.1016/j.jprot.2016.03.006

Cao, Y., Yan, X., Ran, S., Ralph, J., Smith, R. A., Chen, X., et al. (2022). Knockout of the lignin pathway gene *BnF5H* decreases the S/G lignin compositional ratio and improves *Sclerotinia sclerotiorum* resistance in *Brassica napus*. *Plant Cell Environ.* 45 (1), 248–261. doi: 10.1111/pce.14208

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 (1), 222–225. doi: 10.1016/j.molp.2017.09.020

Chen, J., Zhang, H., Deng, S., Du, H., Chen, Z., Zhao, Y., et al. (2021). A backbone parent contributes core genomic architecture to pedigree breeding of early-season indica rice. *J. Genet. Genomics* 48 (11), 1040–1043. doi: 10.1016/j.jgg.2021.07.011

Ding, L. N., Li, T., Guo, X. J., Li, M., Liu, X. Y., Cao, J., et al. (2021). Sclerotinia stem rot resistance in rapeseed: recent progress and future prospects. *J. Agric. Food Chem.* 69 (10), 2965–2978. doi: 10.1021/acs.jafc.0c07351

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 (5), 1255–1270. doi: 10.1111/pbi.13289

Fang, C., Wang, Z., Wang, P., Song, Y., Ahmad, A., Dong, F., et al. (2021). Heterosis derived from nonadditive effects of the *BnFLC* homologs coordinates early flowering and high yield in rapeseed (*Brassica napus l.*). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.798371

Feng, Y., Hu, Y., Fang, P., Zuo, X., Wang, J., Li, J., et al. (2021). Silicon alleviates the disease severity of sclerotinia stem rot in rapeseed. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.721436

Fradgley, N., Gardner, K. A., Cockram, J., Elderfield, J., Hickey, J. M., Howell, P., et al. (2019). A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLoS. Biol.* 17 (2), e3000071. doi: 10.1371/journal.pbio.3000071

Gupta, N. C., Sharma, P., Rao, M., Rai, P. K., and Gupta, A. K. (2020). Evaluation of non-injury inoculation technique for assessing sclerotinia stem rot (*Sclerotinia sclerotiorum*) in oilseed brassica. *J. Microbiol. Meth.* 175, 105983. doi: 10.1016/j.mimet.2020.105983

Hu, H., Tang, Y., Wu, J., Chen, F., Yang, Y., Pan, X., et al. (2021). *Brassica napus* mediator Subunit16 induces BnMED25- and BnWRKY33-activated defense signaling to confer *Sclerotinia sclerotiorum* resistance. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.663536

Jiang, J., Liao, X., Jin, X., Tan, L., Lu, Q., Yuan, C., et al. (2020). *MYB43* in oilseed rape (*Brassica napus*) positively regulates vascular lignification, plant morphology and yield potential but negatively affects resistance to *Sclerotinia sclerotiorum*. *Genes* 11 (5), 581. doi: 10.3390/genes11050581

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 (2), 473–487. doi: 10.1007/s00122-020-03707-9

Kidd, B. N., Edgar, C. I., Kumar, K. K., Aitken, E. A., Schenk, P. M., Manners, J. M., et al. (2009). The mediator complex subunit *PFT1* is a key regulator of jasmonate-dependent defense in *Arabidopsis*. *Plant Cell* 21 (8), 2237–2252. doi: 10.1105/tpc.109.066910

Lai, Z., Schluttenhofer, C. M., Bhide, K., Shreve, J., Thimmapuram, J., Lee, S. Y., et al. (2014). MED18 interaction with distinct transcription factors regulates multiple plant functions. *Nat. Commun.* 5, 3064. doi: 10.1038/ncomms4064

Li, W., Ahn, I. P., Ning, Y., Park, C. H., Zeng, L., Whitehill, J. G., et al. (2012). The U-Box/ARM E3 ligase PUB13 regulates cell death, defense, and flowering time in *Arabidopsis*. *Plant Physiol.* 159 (1), 239–250. doi: 10.1104/pp.111.192617

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 (37), E5123–E5132. doi: 10.1073/pnas.1502160112

Liu, Y., Zhou, X., Yan, M., Wang, P., Wang, H., Xin, Q., et al. (2020). Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, *qGSL-C2*, in rapeseed (*Brassica napus l.*). *Theor. Appl. Genet.* 133 (2), 479–490. doi: 10.1007/s00122-019-03479-x

Li, J., Zhao, Z., Hayward, A., Cheng, H., and Fu, D. (2015). Integration analysis of quantitative trait loci for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Euphytica* 205 (2), 483–489. doi: 10.1007/s10681-015-1417-0

Lyons, R., Iwase, A., Gansewig, T., Sherstnev, A., Duc, C., Barton, G. J., et al. (2013). The RNA-binding protein FPA regulates flg22-triggered defense responses and transcription factor activity by alternative polyadenylation. *Sci. Rep.* 3, 2866. doi: 10.1038/srep02866

Ma, X., Wang, Z., Li, W., Zhang, Y., Zhou, X., Liu, Y., et al. (2019). Resequencing core accessions of a pedigree identifies derivation of genomic segments and key agronomic trait loci during cotton improvement. *Plant Biotechnol. J.* 17 (4), 762–775. doi: 10.1111/pbi.13013

Mei, J., Ding, Y., Lu, K., Wei, D., Liu, Y., Disi, J. O., et al. (2013). Identification of genomic regions involved in resistance against *Sclerotinia sclerotiorum* from wild *Brassica oleracea*. *Theor. Appl. Genet.* 126 (2), 549–556. doi: 10.1007/s00122-012-2000-x

Pressete, C. G., Giannini, L. S. V., de Paula, D. A. C., do Carmo, M. A. V., Assis, D. M., Santos, M. F. C., et al. (2019). *Sclerotinia sclerotiorum* (White mold): cytotoxic, mutagenic, and antimalarial effects *in vivo* and *in vitro*. *J. Food Sci.* 84 (12), 3866–3875. doi: 10.1111/1750-3841.14910

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 (1), 636. doi: 10.1186/s12864-019-5964-y

Shao, Y., Shen, Y., He, F., and Li, Z. (2022). QTL identification for stem fiber, strength and rot resistance in a DH population from an alien introgression of *Brassica napus*. *Plants (Basel)* 11 (3), 373. doi: 10.3390/plants11030373

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 (3), 524–539. doi: 10.1111/tpj.14236

Singh, V., Roy, S., Giri, M. K., Chaturvedi, R., Chowdhury, Z., Shah, J., et al. (2013). *Arabidopsis thaliana FLOWERING LOCUS d* is required for systemic acquired resistance. *Mol. Plant Microbe Interact.* 26 (9), 1079–1088. doi: 10.1094/MPMI-04-13-0096-R

Wang, Z., Bao, L. L., Zhao, F. Y., Tang, M. Q., Chen, T., Li, Y., et al. (2019a). *BnaMPK3* is a key regulator of defense responses to the devastating plant pathogen *Sclerotinia sclerotiorum* in oilseed rape. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00091

Wang, Z., Fang, H., Chen, Y., Chen, K., Li, G., Gu, S., et al. (2014). Overexpression of *BnWRKY33* in oilseed rape enhances resistance to *Sclerotinia sclerotiorum*. *Mol. Plant Pathol.* 15 (7), 677–689. doi: 10.1111/mpp.12123

Wang, Z., Ma, L. Y., Cao, J., Li, Y. L., Ding, L. N., Zhu, K. M., et al. (2019b). Recent advances in mechanisms of plant defense to *Sclerotinia sclerotiorum*. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01314

Wang, Z., Wan, L., Xin, Q., Chen, Y., Zhang, X., Dong, F., et al. (2018). Overexpression of *OsPGIP2* confers *Sclerotinia sclerotiorum* resistance in *Brassica napus* through increased activation of defense mechanisms. *J. Exp. Bot.* 69 (12), 3141–3155. doi: 10.1093/jxb/ery138

Wang, Z., Zhao, F. Y., Tang, M. Q., Chen, T., Bao, L. L., Cao, J., et al. (2020). *BnaMPK6* is a determinant of quantitative disease resistance against *Sclerotinia sclerotiorum* in oilseed rape. *Plant Sci.* 291, 110362. doi: 10.1016/j.plantsci.2019.110362

Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., et al. (2016). Genome-wide association analysis and differential expression analysis of resistance to sclerotinia stem rot in *Brassica napus*. *Plant Biotechnol. J.* 14 (6), 1368–1380. doi: 10.1111/pbi.12501

Wei, D., Mei, J., Fu, Y., Disi, J. O., Li, J., and Qian, W. (2014). Quantitative trait loci analyses for resistance to *Sclerotinia sclerotiorum* and flowering time in *Brassica napus*. *Mol. Breed.* 34 (4), 1797–1804. doi: 10.1007/s11032-014-0139-7

Wu, J., Cai, G., Tu, J., Li, L., Liu, S., Luo, X., et al. (2013). Identification of QTLs for resistance to sclerotinia stem rot and *BnaC.IGMT5.a* as a candidate gene of the major resistant QTL *SRC6* in *Brassica napus*. *PLoS. One* 8 (7), e67740. doi: 10.1371/journal.pone.0067740

Wu, J., Chen, P., Zhao, Q., Cai, G., Hu, Y., Xiang, Y., et al. (2019). Co-Location of QTL for sclerotinia stem rot resistance and flowering time in *Brassica napus*. *Crop J.* 7 (2), 227–237. doi: 10.1016/j.cj.2018.12.007

Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., et al. (2016a). Genome-wide association study identifies new loci for resistance to sclerotinia stem rot in *Brassica napus*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01418

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

Xu, B., Gong, X., Chen, S., Hu, M., Zhang, J., and Peng, Q. (2021). Transcriptome analysis reveals the complex molecular mechanisms of *Brassica napus-Sclerotinia sclerotiorum* interactions. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.716935

Xu, D. F., Li, X. L., Pan, Y. M., Dai, Y. L., Li, P., Chen, F. X., et al. (2014). Genetic diversity and pathogenicity differentiation of *Sclerotinia sclerotiorum* on rapeseed (*Brassica napus l.*) in anhui province, China. *Genet. Mol. Res.* 13 (4), 10704–10713. doi: 10.4238/2014.December.18.12

Yin, X., Yi, B., Chen, W., Zhang, W., Tu, J., Fernando, W. G. D., et al. (2009). Mapping of QTLs detected in a *Brassica napus* DH population for resistance to *Sclerotinia sclerotiorum* in multiple environments. *Euphytica* 173 (1), 25–35. doi: 10.1007/s10681-009-0095-1

Yu, Y., Cai, J., Ma, L., Huang, Z., Wang, Y., Fang, A., et al. (2020). Population structure and aggressiveness of *Sclerotinia sclerotiorum* from rapeseed (*Brassica napus*) in chongqing city. *Plant Dis.* 104 (4), 1201–1206. doi: 10.1094/pdis-07-19-1401-re

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136 (4), 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, F., Huang, J., Tang, M., Cheng, X., Liu, Y., Tong, C., et al. (2019). Syntenic quantitative trait loci and genomic divergence for sclerotinia resistance and flowering time in *Brassica napus*. *J. Integr. Plant Biol.* 61 (1), 75–88. doi: 10.1111/jipb.12754

Zhang, K., Zhuo, C., Wang, Z., Liu, F., Wen, J., Yi, B., et al. (2022). *BnaA03.MKK5-BnaA06.MPK3/BnaC03.MPK3* module positively contributes to *Sclerotinia sclerotiorum* resistance in *Brassica napus*. *Plants (Basel)* 11 (5), 609. doi: 10.3390/plants11050609

Zhao, J., and Meng, J. (2003). Genetic analysis of loci associated with partial resistance to *Sclerotinia sclerotiorum* in rapeseed (*Brassica napus l.*). *Theor. Appl. Genet.* 106 (4), 759–764. doi: 10.1007/s00122-002-1171-2

Zhao, J., and Meng, J. (2008). Detection of loci controlling seed glucosinolate content and their association with sclerotinia resistance in *Brassica napus*. *Plant Breed.* 122 (1), 19–23. doi: 10.1046/j.1439-0523.2003.00784.x

Zhao, J., Udall, J. A., Quijada, P. A., Grau, C. R., Meng, J., and Osborn, T. C. (2006). Quantitative trait loci for resistance to *Sclerotinia sclerotiorum* and its association with a homeologous non-reciprocal transposition in *Brassica napus l.* *Theor. Appl. Genet.* 112 (3), 509–516. doi: 10.1007/s00122-005-0154-5

Zhou, B., Chen, D., Yu, Q., Liu, S., and Yang, J. (1993). Screening and breeding for multiple resistance to sclerotinia stem rot, downy mildew and virus disease. *China J. Oil Crops* 16, 14–17.

Zuo, R., Xie, M., Gao, F., Sumbal, W., Cheng, X., Liu, Y., et al. (2022). The characterization of the *phloem protein 2* gene family associated with resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Int. J. Mol. Sci.* 23 (7), 3934. doi: 10.3390/ijms23073934

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 December 2022
DOI 10.3389/fpls.2022.1081321

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Dengfeng Hong,
Huazhong Agricultural University, China
Mingliang Jiang,
Jilin Agricultural Science and Technology College, China

*CORRESPONDENCE
Zhansheng Li
lizhansheng@caas.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 27 October 2022
ACCEPTED 18 November 2022
PUBLISHED 12 December 2022

CITATION
Yang D, Zhao Y, Liu Y, Han F and Li Z (2022) A high-efficiency PEG-Ca^{2+}-mediated transient transformation system for broccoli protoplasts. *Front. Plant Sci.* 13:1081321. doi: 10.3389/fpls.2022.1081321

A high-efficiency PEG-Ca^{2+}-mediated transient transformation system for broccoli protoplasts

Dongxu Yang†, Yongyu Zhao†, Yumei Liu, Fengqing Han and Zhansheng Li*

Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

Transient transformation of plant protoplasts is an important method for studying gene function, subcellular localization and plant morphological development. In this study, an efficient transient transformation system was established by optimizing the plasmid concentration, PEG4000 mass concentration and genotype selection, key factors that affect transformation efficiency. Meanwhile, an efficient and universal broccoli protoplast isolation system was established. Using 0.5% (w/v) cellulase R-10 and 0.1% (w/v) pectolyase Y-23 to hydrolyze broccoli cotyledons of three different genotypes for 3 h, the yield was more than 5×10^6/mL/g, and the viability was more than 95%, sufficient to meet the high standards for protoplasts to be used in various experiments. The average transformation efficiency of the two plasmid vectors PHG-eGFP and CP507-YFP in broccoli B1 protoplasts were 61.4% and 41.7%, respectively. Using this system, we successfully performed subcellular localization of the products of three target genes (the clubroot resistance gene *CRa* and two key genes regulated by glucosinolates, *Bol029100* and *Bol031350*).The results showed that the products of all three genes were localized in the nucleus. The high-efficiency transient transformation system for broccoli protoplasts constructed in this study makes it possible to reliably acquire high-viability protoplasts in high yield. This research provides important technical support for international frontier research fields such as single-cell sequencing, spatial transcriptomics, plant somatic hybridization, gene function analysis and subcellular localization.

KEYWORDS

broccoli, protoplast, gene function, transient transfection, subcellular localization

Introduction

Broccoli (*Brassica oleracea* L. var. *italica*) is a variety of cruciferous cabbage known as the " nutritional powerhouse" because of its rich content of nutrients such as protein, vitamin C and minerals (Gasper et al., 2005; Fahey et al., 2019; Li et al., 2022). Recent studies have found that broccoli is also rich in sulforaphane (Li Z et al., 2021; Li et al., 2022), an active anticancer compound that significantly reduces the occurrence of various cancers, including liver, lung, stomach, breast, bladder, colorectal, and prostate cancers (Keck et al., 2003; Chilakala et al., 2020; Mahn and Castillo, 2021; Li et al., 2022; Xie et al., 2022). It's reported that sulforaphane could prevent and reduce the risks of cardiovascular and cerebrovascular diseases, hypertension, myopia, depression syndrome, and other conditions (Georgikou et al., 2020). At present, according to the information center of the Ministry of Agriculture, the area in which broccoli is sown in China exceeds 120,000 hectares, accounting for more than 20% of the global output, and China is the main exporter of broccoli (Li et al., ; Li et al., 2019; Huang et al., 2021).

Agrobacterium-mediated genetic transformation technology has been widely used in cruciferous crops, such as oilseed rape (*Brassica napus*) (Bates et al., 2017), mustard (*Brassica juncea* L. Czern.) (Dutta et al., 2008), cabbage (*Brassica oleracea* var. *capitata*) (Brown and Wang, 2004), broccoli (Chen et al., 2001), and Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) (Min et al., 2007), but this method is laborious and yields low transformation efficiency. In addition, there is severe genotype dependence, and self-pollination is required to obtain the T1 generation to observe phenotypes or perform component identification, which increases the time required (Gould et al., 1991; Radke et al., 1992; Wang et al., 2015; Wang Y et al., 2022). Therefore, development of a fast, efficient and genotype-independent genetic transformation system that can speed up the gene function verification of important traits is of great significance. Such a system would also play a significant role in improving the efficiency of mining beneficial genes from the reported genomes of *Brassica* cops. At present, no such system has been reported for important vegetable crops of the *Brassica* genus of the family Cruciferae.

Transient transformation of protoplasts can be used to quickly transfer a large number of target genes for expression and verification in a short period of time (Abel and Theologis, 1994; Duarte et al., 2016). There are many methods for transient transformation, including particle bombardment, microinjection, PEG-mediated transfection, lipofection-mediated transfection and electronic transfection (Wang Y et al., 2022). Among these methods, PEG-mediated transient transformation has the advantages of simple operation, strong universality and stable transformation efficiency (Liu and Friesen, 2012; Sahab et al., 2019a; Wang Y et al., 2022). The establishment of a high-efficiency transient transformation system for broccoli protoplasts, is of great importance for research on gene function, plant physiological and biochemical processes, molecular mechanisms, and some protoplast-based research (single cell, spatial transcriptomics and somatic hybridization) in the cruciferous crops.

In this study, we constructed a high-yielding and high-viability protoplast isolation and purification system for broccoli. Then we took the lead in establishing an efficient and stable transient transformation system for broccoli protoplasts that can be used for rapid analysis and validation of gene functions. Based on this system, we successfully determined the subcellular localization of the gene product of the enhanced rhizoctonia-resistance gene *CRa* (Ueno et al., 2012) of *Brassica* and those of two key glucosinolate regulatory genes, *Bol029100* and *Bol031350* (Rubel et al., 2020).

Results

Isolation of protoplasts from broccoli cotyledons

A high-efficiency protoplast isolation system was successfully constructed using broccoli explants of various genotypes. To determine the universality and efficiency of the system, three genotypes of broccoli protoplasts were all isolated by the same method. The cotyledons of sterile seedlings that had been cultured for 7-10 days were preincubated in CM solution (Table S1) for 12 h, and strips of approximately 1 mm were then cut from the leaves and placed in 0.3 M mannitol enzymatic hydrolysis solution for enzymatic hydrolysis. The process did not require a vacuum environment and involve toxic reagents, which indicated it was simple and reliable in operation. The leaf material was hydrolyzed with 0.5% (w/v) cellulase R-10 and 0.1% (w/v) pectolyase Y-23 on a shaking table for 3 h. Using this method, the yield of B1, B40, and B42 protoplasts all reached or exceeded 5.05×10^6 protoplasts/g. The B1 genotype yielded 5.33×10^6-7.05×10^6 protoplasts/g, the B42 genotype yielded 6.23×10^6-9.84×10^6 protoplasts/g, and the yield from the B40 genotype was 5.05×10^6-6.04×10^6 protoplasts/g. These yields met the standard needed for protoplast transformation experiments (1.0×10^6-2.0×10^6 protoplasts/mL). At the same time, the protoplast viability was determined by fluorescein diacetate (FDA) to be greater than 95% (Figure 1) (Table S2). Microscope observation showed that the purified protoplasts were relatively unfragmented and that the preparations contained little cell debris. Thus, they fully met the standard required for various protoplast-related experiments. From this result, protoplasts with high viability were obtained in high yield from the cotyledons of broccoli plants of the three

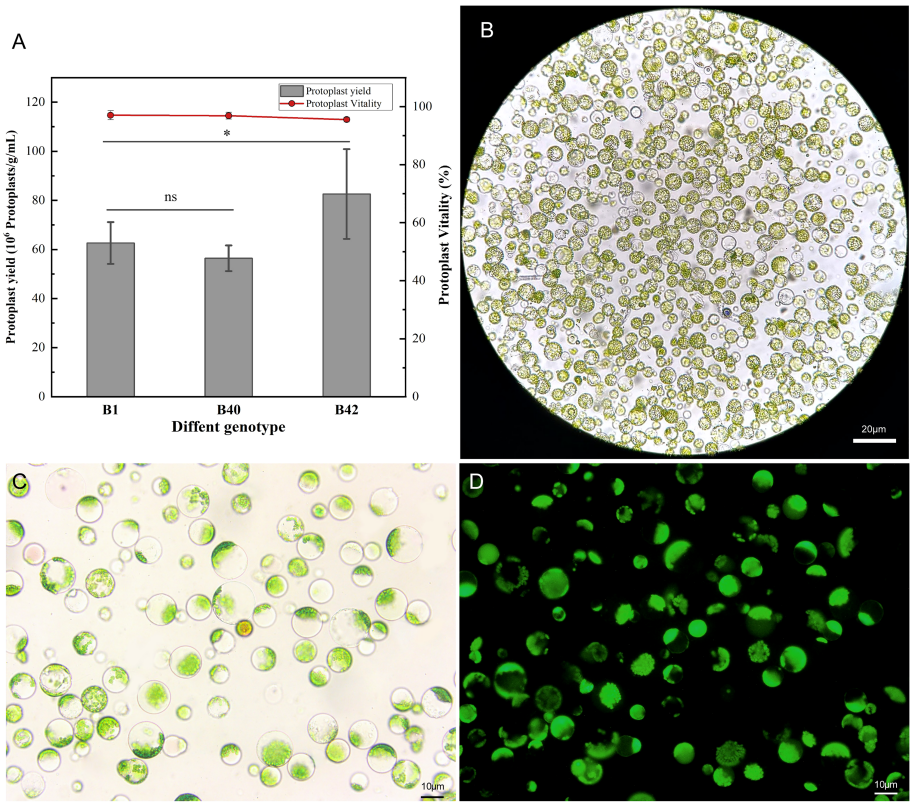


FIGURE 1
Yield and viability of cotyledon protoplasts of different broccoli genotypes. **(A)** Statistics on cotyledon protoplast yield and viability of three genotypes of broccoli, B1, B40 and B42. Data represent the means of 3 values, and error bars show the standard error values; Statistical significance levels (Student's t-test) are shown: ns, no significant; *$p < 0.05$. **(B)** Microscopic examination of B1 genotype protoplast, scale bar = 20 μm. **(C)** The bright field images of B1 genotype protoplast for fluorescein diacetate (FDA) microplate assay, scale bar = 10 μm. **(D)** The dark field images of B1 genotype protoplast for fluorescein diacetate (FDA) microplate assay, scale bar = 10 μm.

genotypes, and there were no significant differences between the B1 with the B40 in yield or viability ($p < 0.05$), demonstrating the universality and efficiency of the method.

Construction of a transient transformation system for broccoli protoplasts

The PHG-eGFP (12.7 kb) and CP507-YFP (10.7 kb) plasmids were transferred into the prepared B1 protoplasts, and the protoplasts were cultured at 25°C for 13-16 h in the dark. The green fluorescence signal of GFP in protoplasts transfected with PHG-eGFP and the yellow fluorescence of YFP in protoplasts transfected with CP507-YFP were observed under excitation at 488 nm (Figure 2A). The result showed that the plasmid DNA that encoded the fluorescent protein could be transferred into broccoli protoplasts and the fluorescent proteins were stably expressed. The two plasmids were also transformed into B40 and B42 protoplasts, and the fluorescent proteins were also stably expressed in those protoplasts. Thus, a transient transformation system for broccoli protoplasts was successfully constructed in this study.

Effect of plasmid concentration on protoplast transformation efficiency

A total of 5 μg, 10 μg, and 15 μg plasmid vectors of PHG-eGFP and CP507-YFP were transferred into 200 μL (approximately 4×10^5) protoplasts obtained from broccoli genotypes, and the resulting transformation efficiency was calculated shown in Figures 2B, C. Although the average efficiency of transformation by PHG-eGFP and CP507-YFP increased slightly as the plasmid concentration increased, there was no significant difference between treatments in any of the three genotypes ($p < 0.05$). When the plasmid mass was increased stepwise from 5 μg to 15 μg, the efficiency of transformation of B1 protoplasts by PHG-eGFP was respectively 58.6-65.8%, 59.5-64.7%, and 59.4-73.1%, with

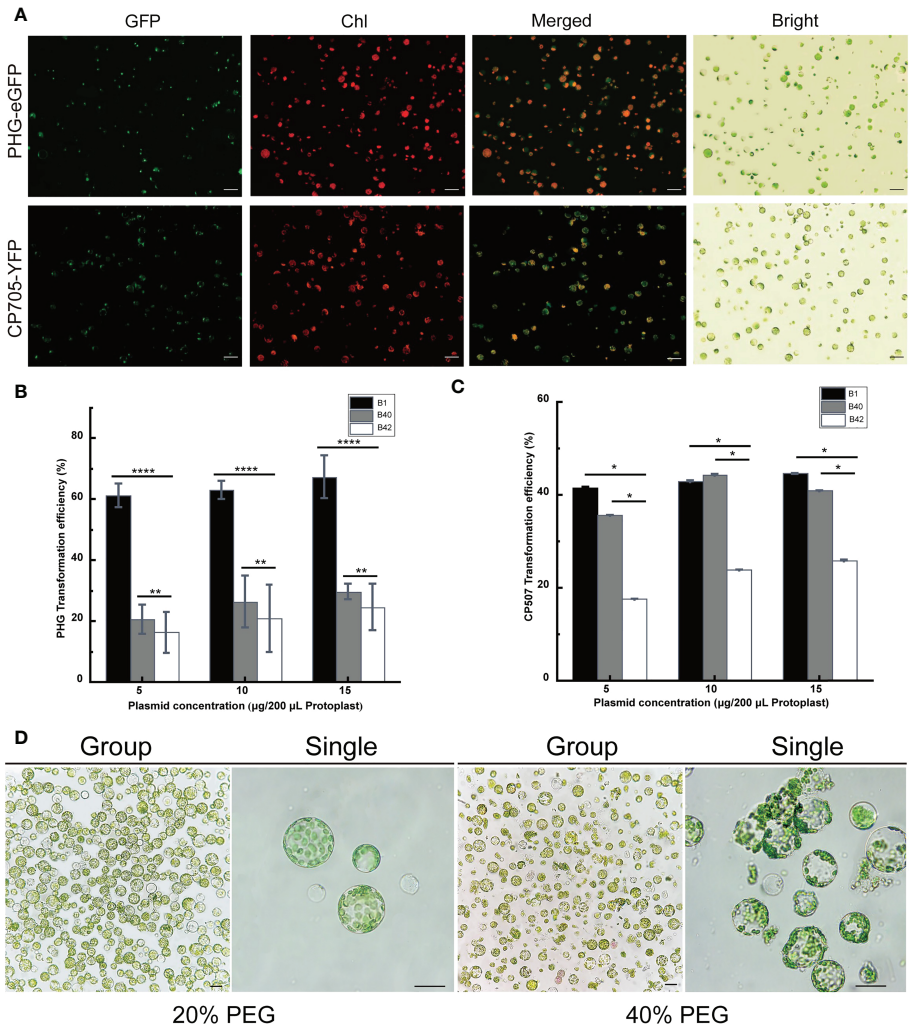


FIGURE 2
Optimization of the transient transformation system for broccoli protoplasts. Transient expression of different constructs in B1 protoplasts. Samples were visualized under a fluorescence microscope. **(A)** The 35S::eGFP PHG-eGFP and 35S::YFP CP507-YFP were transiently expressed in B1 cotyledon protoplasts. Merged images of GFP or YFP and chlorophyll autofluorescence (Chl) as well as bright field images of protoplasts were shown. Scale bar = 20 μm. **(B, C)** Comparison of transfection efficiency of 5-15 μg/200 μL protoplasts PHG-eGFP and CP507-YFP plasmids in B1, B40 and B42 cotyledon protoplasts, expressed as the ratio of GFP-positive cells to the total number of protoplasts (n≥100). Data represent the means of 3 values, and error bars show the standard error values; Statistical significance levels (Student's t-test) are shown: ns, no significant; *$p < 0.05$; **$p < 0.01$; ****$p < 0.0001$. **(D)** After 20% PEG4000 and 40% PEG4000 induction treatment for 20 min, the protoplasts were incubated in W5 solution for 12 h to observe the protoplast culture by microscopy. Group images and single images were observed under 10× and 40× microscope, respectively. Scale bar = 10 μm.

mean values of 61.4%, 63.0%, and 67.4%, respectively. For B40, they were respectively 15.5-25.0%, 17.6-34.5%, and 27.4-32.4% with mean values of 20.6%, 26.6%, and 29.7%. And in B42, they were 12.8-24.5%, 10.4-32.2%, and 16.8-30.4% with mean values of 16.3%, 21.4%, and 24.7%, respectively. When the same amounts of the CP507-YFP plasmid were used, the efficiency of transformation of B1 were respectively 42.2-44.8%, 30.0-55.2%, and 38.2-48.8% with mean values of 41.7%, 43.0%, and 44.7%; the efficiency for B40 were respectively 25.4-41.8%, 35.0-53.0%, and 38.8%-44.2% with mean values of 35.7%, 44.4%, and 41.0%. While the efficiencies for B42 were respectively 15.4-19.8%, 21.2-27.4%, and 20.0-33.3% with mean values of 17.7%, 24.0%, and 26.0%. The results indicated that over the range of 5-15 μg plasmid/4×10^5 protoplasts, the amount of plasmid had little or nearly no effect on the transfection efficiency. The transformation efficiency of PHG-eGFP in B1 varied the most with increasing plasmid concentration (58.6%-73.1%). The average

transformation efficiency of the other two plasmids were less than 9.1% (Figures 2B, C) (Table S3).

Effect of PEG4000 mass concentration on cotyledon protoplast transformation efficiency

Totally, 5 μg PHG-eGFP vectors were added to 200 μL of B1, B40 and B42 cotyledon protoplasts, and transfection was induced by the addition of an equal volume of 20% (w/v) PEG4000 or 40% (w/v) PEG4000 PEG-Ca^{2+} and incubation for 20 min. The results showed that the efficiency of transformation using 40% PEG4000 was higher than 20% PEG4000, but when the protoplasts were cultured in the presence of 40% PEG4000, a large number of the protoplasts ruptured, and this seriously affected the subsequent observation and culture of the protoplasts (Figure 2D). Similarly, the protoplasts transformed by CP507-YFP in the presence of 40% PEG-Ca^{2+} showed obvious rupture. Thus, this study found that the optimal mass concentration of PEG4000 for cotyledon protoplast transformation of the three broccoli genotypes was 20%.

Transformation efficiency of different broccoli genotypes

The transformation efficiency of protoplasts for B1, B40 and B42 genotypes were compared under the conditions of 5 μg plasmid concentration and 20% PEG-Ca^{2+} induction solution. From Figure 2A and Table S3, we could find that the PHG-eGFP and CP507-YFP plasmid vectors all yielded higher transformation efficiency in B1 protoplasts, and that there were significant differences in the transformation efficiency obtained using different plasmid vectors ($p < 0.05$) to 65.8% with an average of 61.4%. And for CP507-YFP, transformation efficiency ranged from 38.2% to 45.0% with an average of 41.7%. The transformation efficiency of PHG-eGFP and CP507-YFP were both in the lowest level in B42 protoplasts, which respectively ranged from 12.5% to 24.8% and from 15.0% to 20.6% with the average values were 16.3% and 17.7%. So, our research had shown that the efficiency of protoplast transformation was obviously influenced by recipient genotypes and by vector types.

Subcellular localization of target genes

The expression of three target genes, *CRa*, *Bol021900* and *Bol031350*, linked to GFP in B1 protoplasts cultured in W5 suspension was verified by laser confocal microscopy. It has been reported that *Bol021900* and *Bol031350* are involved in the regulation of sulforaphane metabolism and that *CRa* expression significantly enhances resistance to clubroot in *Brassica* crops (Ueno et al., 2012; Ce et al., 2021). After induction with 20% PEG-Ca^{2+}, the protoplasts were cultured in W5 solution for 12-15 h. Under microscopic observation, green fluorescence was observed in the protoplasts, and the three gene products and their fusion proteins were all located in the nucleus (Figure 3). The expression of fluorescent proteins in broccoli protoplasts demonstrated the validity and stability of this transformation system, which showed that it can be successfully applied to study the subcellular localizations of the products of specific genes.

Discussion

To date, protoplasts have been successfully isolated from approximately 40 families, more than 100 genera, and more than 400 species of plants (Reed and Bargmann, 2021). The protoplasts have been successfully isolated from the family Cruciferae focus on *Arabidopsis. thaliana* (Yoo et al., 2007; Ryu et al., 2019), and there are a few reports on *Brassica napus* (Kang et al., 2017; Li X et al., 2021), Chinese cabbage (Sivanandhan et al., 2021), cabbage (Fu et al., 1985) and cauliflower(*Brassica. oleracea* var. *botrytis*) (Yang et al., 1994), but some common problems such as low protoplast yield and viability still be improved. Despite the fact that there are a few reports on transient transformation systems based on protoplasts in Cruciferae, the problems of low transformation efficiency and genotype dependence restrict the wide application

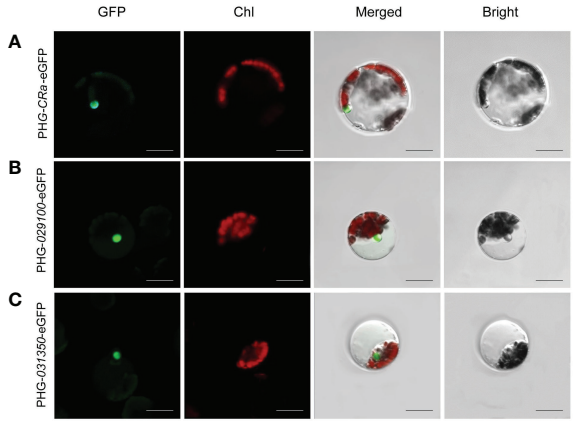


FIGURE 3
Subcellular localization of *CRa*, *Bol021900* and *Bol031350* genes. **(A-C)** presented PHG-*CRa*-eGFP, PHG-*029100*-eGFP and PHG-*031350*-eGFP were all located in the nucleus and transiently co-expressed in broccoli B1 protoplasts; Individual and merged images of GFP and chlorophyll autofluorescence (Chl) as well as bright field images of protoplasts were shown. Scale bars = 10 μm.

of this technology (Sun et al., 2019; Sahab et al., 2019b). At present, plant protoplast-based technologies are being widely used in frontier fields such as single-cell sequencing (Armand et al., 2021; Marand et al., 2021), subcellular localization (Mackon et al., 2021), somatic cell hybridization (Xia, 2009), and plant morphogenesis (Yu et al., 2020). The establishment of an efficient and universal protoplast isolation system is the experimental premise and the necessary support for research in the above frontier fields.

In this study, a genotype-independent high-efficiency protoplast isolation and purification technology system for broccoli was successfully constructed. The yield and vitality of protoplasts obtained by this method are sufficient to meet the needs for transformation and various related experiments based on protoplasts. The coordination of enzymatic hydrolysis time with enzyme hydrolysate concentration ratio is a key factor in obtaining high protoplast yield and high viability (Liqing et al., 2005; Cao et al., 2014; Beard et al., 2021). In this study, 0.5% cellulase and 0.1% pectinase were used to isolate protoplasts from the cotyledons of broccoli of three different genotypes; the use of an enzymatic hydrolysis time of 3 h ensured high yield (> 5×10^6/g) and high vitality (> 95%) of the resulting protoplasts. The broccoli protoplast isolation and purification system constructed in this study is superior to the systems previously reported for other cruciferous crops such as Chinese cabbage (Sivanandhan et al., 2021), cabbage (Kiełkowska and Adamus, 2019), rape (Li X et al., 2021), cauliflower (Yang et al., 1994) and others. In addition, we observed that broccoli genotype is not a key factor affecting protoplast yield and viability. Similar protoplast yields and viabilities were obtained for the three tested genotypes of broccoli under the same enzymatic hydrolysis conditions, indicating that the system is universal to broccoli materials of different genotypes. This finding provides an important scientific basis for improving the yield and viability of protoplasts isolated from other cruciferous crops.

Transient transformation system for protoplast mediated by PEG-Ca^{2+} have been successfully constructed and applied in field crops such as maize(*Zea mays* L.) (Cao et al., 2014; Hu et al., 2020), wheat(*Triticum aestivum* L.) (Luo et al., 2022), rice(*Oryza sativa* L.) (Page et al., 2019) and cotton(*Gossypium* spp) (Wang P et al., 2022), but the use of such systems in cruciferous crops has been limited to *A. thaliana*. There are several reports on cabbage and Chinese cabbage, but the transformation efficiency of those two species is less than 45%, and the transient transformation system displays serious genotype-dependent performance (Sun et al., 2019; Sahab et al., 2019b). The PEG-Ca^{2+}-mediated protoplast transient transformation system established in this study was also genotype-dependent, and the transformation efficiency of the three broccoli genotypes differed significantly ($p < 0.05$). The transformation efficiency of the three plasmids in the inbred broccoli line of B1 genotype varied widely from 39.7.6% to 89.2%, and the efficiency was only 15.5% to 53.8% in the B40 genotype and 10.9% to 45.2% in the B42 genotype. It verifies that genotype is an important factor that affects the instantaneous transformation efficiency of *Brassica* crops such as broccoli (Sun et al., 2019; Sahab et al., 2019b). At the same time, our study found that a higher transient transformation efficiency can be obtained in broccoli genotype B1 mediated by *Agrobacterium*, which suggested that genotype good for genetic transformation can ensure a rapid acquisition of its genetically transformed progeny (Han et al., 2021).

We also optimized two other key factors that affect broccoli protoplast transformation: plasmid concentration and mass concentration of PEG4000. In this study, we found no significant difference ($p < 0.05$) in transformation efficiency in the same broccoli genotype at plasmid concentrations of 5-15 μg, consistent with the results of previous studies (Zhang et al., 2011). The optimal mass concentration of PEG4000 for inducing transient transformation of broccoli protoplasts was 20%, not the 40% PEG4000 concentration that is used in most plant protoplasts (Yoo et al., 2007; Zhang et al., 2011; Cao et al., 2014; Biswas et al., 2022). At a PEG4000 concentration of 40%, the transformation efficiency of broccoli protoplasts was somewhat improved, but protoplast rupture was observed after culture of the material; this is a novel finding. In conclusion, genotype and the specific plasmid used both had significant effects on the efficiency of transformation of cyanobacterial protoplasts. Thus, to improve protoplast transformation efficiency, we should first select genotypic materials that have the potential for high transformation efficiency and then choose an appropriate PEG4000 induction concentration. The mechanism by which genotype affects the transformation efficiency of protoplasts remains to be further investigated.

Finally, we studied the subcellular localization of three target genes, *Bol029100*, *Bol031350* and *CRa*. And the *Bol029100* and *Bol031350* genes encode FMO $_{GS\text{-}OX5}$ (flavin-containing monooxygenase, FMO $_{GS\text{-}OX5}$); broccoli contains two homologous copies of this gene (Wang et al., 2011; Zhao et al., 2021), which is present in only one copy in *Arabidopsis* (*AT1G12140*) (Kong et al., 2016). The gene encodes flavin monooxygenase (FMO $_{GS\text{-}OX}$), which catalyzes the oxidation of sulfur atoms in methylthioalkyl mustard oleosides (MT GSLs) to produce the anticancer active precursor of sulforaphane (SFN, SF), 4-methyl sulfinyl butylthioside (glucoraphanin, GRA) (Lee et al., 2017; Wang et al., 2017; Tao et al., 2022). While we found that two homologous copies of the gene are present in broccoli, and the functional differences between these genes need to be further verified. This study provides a scientific basis for revealing their functional properties and investigating their expression. Subcellular localization of the protein encoded by the clubroot resistance gene *CRa*, was firstly reported in broccoli, which might provide more evidence of clubroot resistant in *Brassica oleracea* crops.

In this study, an efficient isolation and transient transformation system for broccoli protoplasts were constructed. This system makes possible the acquisition of high-yield and high-viability

protoplasts and provides important technical support for research in frontier fields such as single-cell sequencing, spatial transcription analysis, somatic cell hybridization, gene function analysis, and subcellular localization. Using this transformation system, T1-generation transformants can be directly obtained, facilitating the observation and analysis of their development from single cells to plants. Moreover, this system reduces the culture time needed to obtain positive plants, greatly enhancing the transformation efficiency. Above all, it provides new ways and new ideas for histological analysis, gene function verification and morphological construction (Figure 4).

Conclusions

We constructed an efficient protoplast isolation and purification system for broccoli that is universal for different genotypes and can be used to obtain protoplasts with high viability in high yield, thus overcoming the limitation of genotype dependence. A PEG-Ca^{2+}-mediated efficient transient transformation technology system for broccoli protoplasts was optimized and established. The transformation efficiency it yielded was significantly higher than those provided by systems previously developed for cruciferous crops and showed significant enhancement compared with systems for non-cruciferous crops. A new finding of this study is that the use of 20% PEG4000 during induction yielded significantly better than 40% PEG4000 in broccoli, which better ensured the integrity of the resulting protoplasts. In summary, the efficient transient transformation system for broccoli protoplasts can be successfully applied to subcellular localization studies. The system provides a new method and technical support for protoplast isolation, subcellular localization studies and research on gene function, plant physiological and biochemical processes, molecular mechanisms in cruciferous crops such as broccoli.

Materials and methods

Materials and reagents

The cotyledons of sterile seedlings of various genotypes of the broccoli inbred lines B1, B40 and B42 were used as exosomes for protoplast isolation (Figure 5). All materials were cultivated by Institute of Vegetable and Flower Research, Chinese Academy of Agricultural Sciences (IVF-CAAS). B1 was the F_6 generation of the inbred line material, and it exhibited medium-early maturity, semierect plant type, dome-shaped head, blue–green buds, medium fine flower buds, and no lateral branching in the field (Figure 5A). B40 was the F_6 generation of the inbred line material, and it showed medium-late maturity, erect plant type, semidome-shaped head, dark green buds, medium flower buds and no lateral branching in the field (Figure 5B). B42, the F_7 generation of inbred line material, exhibited early maturity, semierect plant shape, dome-shaped head, green bulbs, fine buds and few lateral branches in the field (Figure 5C).

An eGFP expression cassette was created using the plasmids PHG-eGFP (12.7 kb), which encode green fluorescent protein and promote eGFP expression using CaMV 35S promoter and rbcS ployA as a terminator. The plasmid CP507-YFP (10.7 kb) containing yellow fluorescent protein uses CaMV 35S promoter to promote YFP expression and NOS as the terminator to form a YFP expression box (Figures 6A, B). The CP507-YFP plasmid was provided by the Quality Molecular Research Group of IVF-CAAS, and was propagated and preserved in our laboratory. The plasmid was amplified in the *E. coli* DH5α strain, and the EndoFree Maxi plasmid kit (TIANGEN Biotech Co., Ltd., Beijing, China) was used to extract the plasmid. The purified plasmid was stored at –20°C for later use.

Cellulase "onozuka" R-10 and pectolyase Y-23 were purchased from YAKULT HONSHA CO., LTD. Mannitol, bovine serum albumin (BSA), fluorescein diacetate (FDA), 2-(N-morpholino) ethanesulfonic acid (MES), polyethylene glycol 4000 (PEG 4000) and other reagents were purchased from Beijing Exelon Biotechnology Co., Ltd. The *E. coli* DH5α strain was purchased from Shanghai Weidi Biotechnology Co., Ltd.

Cultivation of aseptic seedlings

Selected filled seeds were sterilized in 50 mL centrifuge tubes containing 75% alcohol for 3 min and then poured into 8% sodium hypochlorite for 8 min, rinsed three times with sterile water, and spread on seeding medium (4.45 g/L MS + 28 g/L sucrose + 8 g/L agar, pH 5.8). The cultures were incubated for 7-10 d at 25°C in low light (150 μmol m-2 s-1, 16 h light, 8 h dark).

Protoplast isolation and purification

Upper fully expanded broccoli cotyledons of aseptic seedlings cultured for 7-10 d were selected (Figure 5), spread over the bottoms of sterile 9-cm glass Petri dishes filled with CM (Table S1), and incubated in the dark at 4°C for 12 h. After pretreatment, the CM was discarded, a scalpel was used to make a scratch approximately 1 mm long on the back of the leaf, and 10 mL of preconfigured enzyme digestion solution (0.5% cellulase R-10, 0.1% pectolyase Y-23, and 0.6 M mannitol, pH 5.8) was added. The material was placed on a shaker and enzymatically digested at 25°C in the dark for 3 h. After enzymatic digestion, the dishes were gently shaken in a single

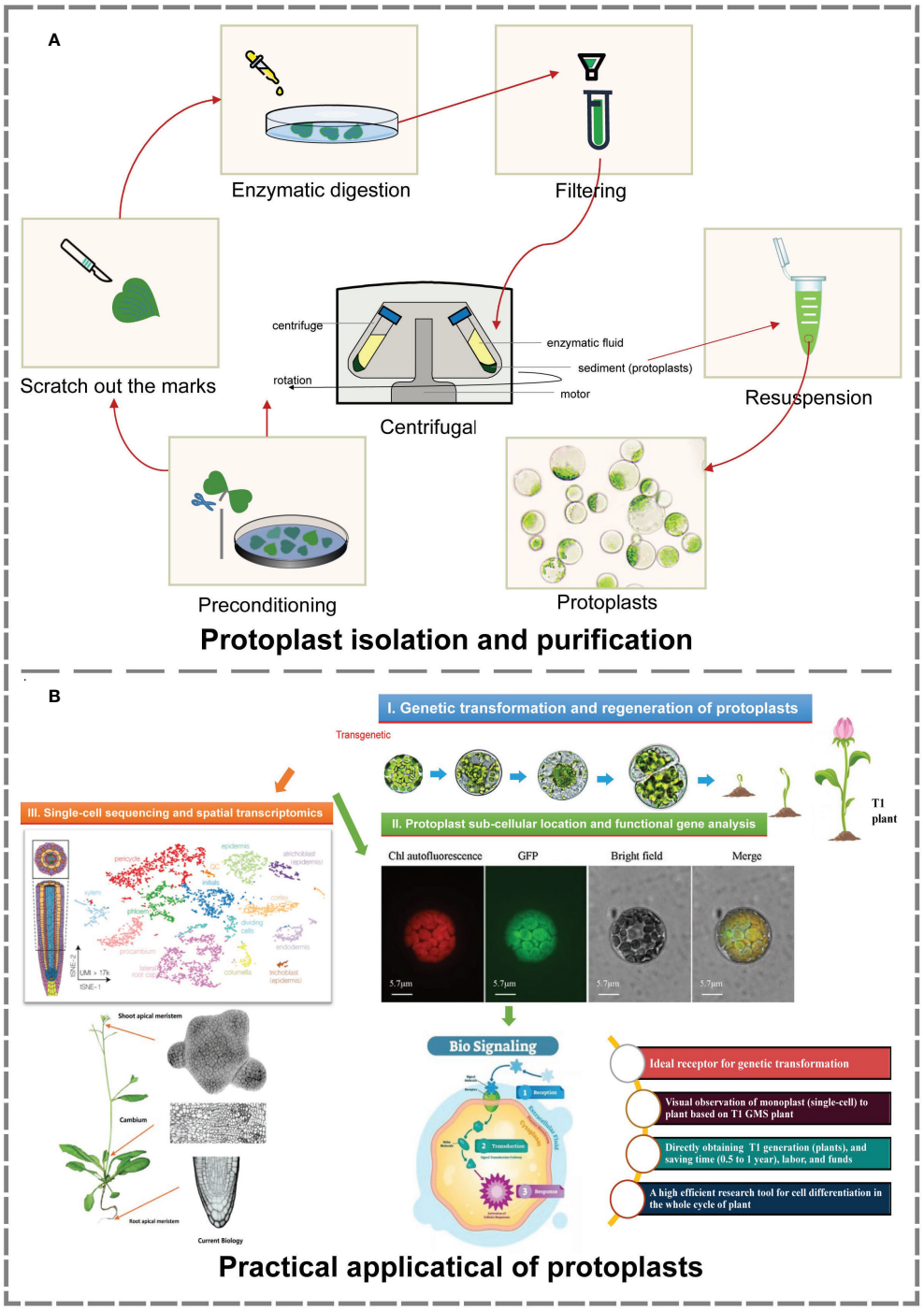


FIGURE 4
Isolation, purification and practical application of protoplasts (Buckley, 2015; Greb and Lohmann, 2016; Tanveer and Yousaf, 2020; Moses and Pachter, 2022). **(A)** The process of protoplast isolation and purification. The material and production of images based on our research. **(B)** Main practical applications of protoplasts. Image material from references, typesetting and production based on our research.

direction to release the protoplasts. The leaf residue was removed by passage of the material through a 50.0-μm nylon membrane sieve. The material that passed through the sieve was collected in a 10-mL round-bottom centrifuge tube. After centrifugation of this material at 700 r/min for 5 min, the supernatant was discarded, 7.0 mL of W5 (154 mM NaCl, 125.0 mM $CaCl_2$, 5.0 mM KCl and 2.0 mM MES, pH 5.7) was added to the pellet, the resulting suspension was centrifuged at

10.3389/fpls.2022.1081321

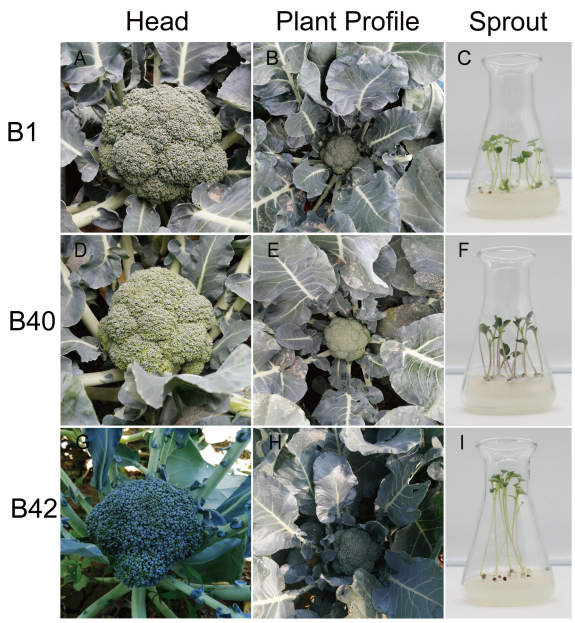


FIGURE 5
The profiles of broccoli plants and sprouts. **(A–C)** Head and plant profiles of B1 genotype broccoli plant. Representative healthy 10-day-old B1 broccoli sprout on seeding medium used for protoplast isolation. **(D–F)** Head and plant profiles of B40 genotype broccoli plant. Representative healthy 10-day-old B40 broccoli sprout on seeding medium used for protoplast isolation. **(G–I)** Head and plant profiles image of B42 genotype broccoli plant. Representative healthy 10-day-old B42 broccoli sprout on seeding medium used for protoplast isolation.

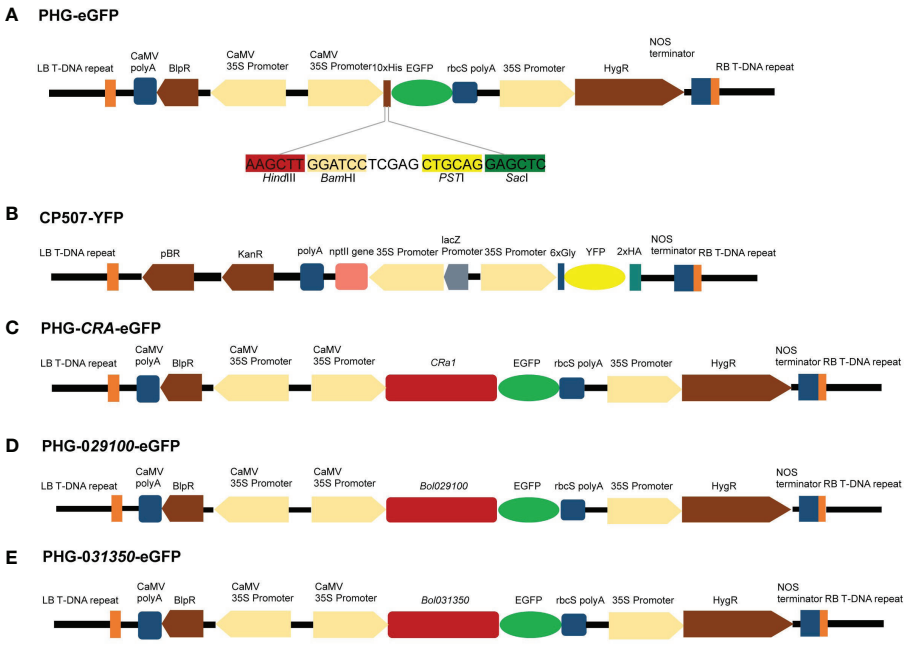


FIGURE 6
Maps of vectors with target genes. **(A)** Genetic map of PHG-eGFP and information of four enzymatic cut sites. **(B)** Genetic map of CP507-YFP. **(C)** Genetic map of PHG-*CRa*-eGFP. **(D)** Genetic map of PHG-*029100*-eGFP. **(E)** Genetic map of PHG-*031350*-eGFP.

700 r/min for 3 min, and the supernatant was again discarded. The above process was repeated twice. The protoplasts were then immediately resuspended in MMG solution (0.4 M mannitol, 15.0 mM $MgCl_2$ and 4.0 mM MES, pH 5.7) at a density of 1.0-2.0×10^6 protoplasts mL^{-1}.

The yield of protoplasts was calculated using the hemocytometer counting method. The yield of obtained protoplasts (cells g^{-1}) = N×10^4×V×m^{-1}, where N = the number of protoplasts counted in a hemocytometer chamber, V = the volume of diluted protoplasts, and m = the fresh weight of the cotyledons used for protoplast isolation. The viability of the protoplasts obtained was checked using fluorescein diacetate (FDA). Ten microliters of 0.01% fluorescein diacetate (FDA) were added to 500 microliters of protoplast suspension. After 5 min, the protoplasts were examined with an Olympus BX51 fluorescence microscope (green fluorescence, Olympus, Japan). The viability of the obtained protoplasts (%) = the number of protoplasts with green fluorescence/total protoplasts in the field (×100%). The experiment was repeated three times, and the average value was taken (n = 3).

The isolation and purification of broccoli cotyledon protoplasts of the three genotypes was performed using the above procedure.

PEG-Ca^{2+}-mediated protoplast transfection

The method referenced by Zhang et al. (2011) to transfect broccoli protoplasts was slightly optimized (Zhang et al., 2011). Plasmids (5-15 μg) were added to 200 μL of protoplast suspension (approximately 4×10^5 protoplasts), followed by addition of an equal volume of PEG-Ca^{2+} solution (PEG4000, 0.2 M mannitol and 0.1 M $CaCl_2$), and transformation was allowed to proceed in the dark at room temperature (approximately 25°C) for 20 min. Two volumes of W5 solution were then added to stop the reaction, the sample was centrifuged at 700 r/min for 3 min, the supernatant was discarded, and 1.0 mL of W5 solution was added to resuspend the pelleted material. The protoplasts were transferred to a 3.0-cm petri dish and cultured at 25°C in the dark for 12-15 h. The concentration of protoplasts used for transformation and the amount of W5 solution used to culture the transformants can be appropriately increased or reduced according to the purpose of the experiment. At the end of the culture period, the expression of fluorescent protein was observed under a fluorescence microscope, and the results were used to calculate the transformation efficiency of different materials.

Protoplast transformation efficiency was defined as the number of protoplasts with green fluorescence/total protoplasts in the field (×100%). The experiment was repeated three times, and the average value was taken (n = 3).

To obtain higher protoplast transformation efficiency, stepwise gradients of the factors affecting the transformation were tested: plasmid masses of 5, 10 and 15 μg and mass concentrations of PEG4000 in the PEG-Ca^{2+} solution of 20% (w/v) and 40% (w/v) were tested. The transformed materials were cotyledon protoplasts of different broccoli genotypes of B1, B40, and B42.

Plasmids of PHG-*CRa*-eGFP, PHG-*029100*-eGFP and PHG-*031350*-eGFP

To construct the PHB-CRa plasmid, the full-length 4.6-kb *CRa* gene (derived from cabbage) cloned in our laboratory was ligated to the PHB plasmid after single digestion of the PHB plasmid with XhoI. Subsequently, PHB-CRa was digested with BamHI/XbaI, and DNA ligase (M0202M, NEB, MA, USA) was used to ligate eGFP to construct a PHG-*CRa*-eGFP expression vector (Figure 6C).

The *Bol029100* gene was amplified by PCR using the specific primers P1 (F: ATCCTCGAGCTGCAGG AGCTCATG GCACCCGCATGCAGCAGAATC; R: CCCTTGCTCA CCATCACTAGTCCAATCAGCTTCTGTCTCAAGAAATC), and the *Bol031350* gene was amplified using the specific primers P2 (F: ATCCTCGAGCTGCAGGAGCTCATGGC ACCCTCTTGCAGTCCAATC; R: CCCTTGCTCACC ATCACTAGTCCCCCAAATATCTGAGAGGGAATAAC). The PCR fragment was recovered after gel electrophoresis and ligated to the digested plasmid PHG-eGFP to construct the PHG-*029100*-eGFP and PHG-*031350*-eGFP expression vectors (Figures 6D, E).

Microscopy

Fluorescence microscopy (Olympus BX51, Japan) was used to observe and photograph the expression of green and red fluorescent proteins in broccoli protoplasts transfected with the PHG-eGFP, and CP507-YFP plasmids. The subcellular localization of the fluorescent proteins was observed through laser confocal microscopy (Olympus BX51, Japan). GFP was excited using a 488-nm laser line.

Statistical analysis

All data were analyzed by analysis of variance (ANOVA) using SPSS Statistics version 19.0 (IBM®, Chicago, IL, USA). Student's t-test or one-way analysis of variance (ANOVA), and Duncan's multiple tests were used to assess significant differences ($p < 0.05$).

Data availability statement

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

Author contributions

ZL: Conceptualization, Formal analysis, Project administration, Resources, Writing - review & editing, Validation. DY: Data curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. YZ: Methodology, Validation, Visualization, Software, Investigation. YL and FH: Methodology, Review & editing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (32172580), the Chinese Agricultural Research System (CARS-23-A05), the Basic Research Fund of the Central Public Welfare Scientific Research Institution (IVF-BRF2021003), the Agricultural Science and Technology Innovation Program (ASTIP) and the State Key Laboratory of Vegetable Germplasm Innovation (SKL-VGI) Research grants.

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.2022.1081321/full#supplementary-material

References

Abel, S., and Theologis, A. (1994). Transient transformation of arabidopsis leaf protoplasts: a versatile experimental system to study gene expression. *Plant J.* 5, 421–427. doi: 10.1111/j.1365-313x.1994.00421.x

Armand, E. J., Li, J., Xie, F., Luo, C., and Mukamel, E. A. (2021). Single-cell sequencing of brain cell transcriptomes and epigenomes. *Neuron* 109, 11–26. doi: 10.1016/j.neuron.2020.12.010

Bates, R., Craze, M., and Wallington, E. J. (2017). Agrobacterium-mediated transformation of oilseed rape (Brassica napus). *Curr. Protoc. Plant Biol.* 2, 287–298. doi: 10.1002/cppb.20060

Beard, K. M., Boling, A. W. H., and Bargmann, B. O. R. (2021). Protoplast isolation, transient transformation, and flow-cytometric analysis of reporter-gene activation in cannabis sativa l. *Ind. Crops Products* 164, 113360. doi: 10.1016/j.indcrop.2021.113360

Biswas, S., Wahl, N. J., Thomson, M. J., Cason, J. M., McCutchen, B. F., Septiningsih, E. M, et al. (2022). Optimization of protoplast isolation and transformation for a pilot study of genome editing in peanut by targeting the allergen gene ara h 2. *Int. J. Mol. Sci.* 23(2), 837. doi: 10.3390/ijms23020837

Brown, D. C. W., and Wang, H. Y. (2004). "Agrobacterium-mediated transformation of cabbage in transgenic crops of the world," in *Essential protocols*. Ed. I. S. Curtis (Netherlands: Springer), 361–378.

Buckley, T. N. (2015). The contributions of apoplastic, symplastic and gas phase pathways for water transport outside the bundle sheath in leaves. *Plant Cell Environ.* 38, 7–22. doi: 10.1111/pce.12372

Cao, J., Yao, D., Lin, F., and Jiang, M. (2014). PEG-mediated transient gene expression and silencing system in maize mesophyll protoplasts: a valuable tool for signal transduction study in maize. *Acta Physiologiae Plantarum* 36, 1271–1281. doi: 10.1007/s11738-014-1508-x

Ce, F., Mei, J., He, H., Zhao, Y., Hu, W., Yu, F., et al. (2021). Identification of candidate genes for clubroot-resistance in brassica oleracea using quantitative trait loci-sequencing. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.703520

Chen, L.-F. O., Hwang, J.-Y., Charng, Y.-Y., Sun, C.-W., and Yang, S.-F. (2001). Transformation of broccoli (Brassica oleracea var. italica) with isopentenyltransferase gene *via* agrobacterium tumefaciens for post-harvest yellowing retardation. *Mol. Breed.* 7, 243–257. doi: 10.1023/A:1011357320259

Chilakala, R. R., Manchikalapudi, A. L., Kumar, A., and Sunkaria, A. (2020). Sulforaphane attenuates a beta oligomers mediated decrease in phagocytic activity of microglial cells. *Neuroscience* 429, 225–234. doi: 10.1016/j.neuroscience.2020.01.002

Duarte, P., Ribeiro, D., Carqueijeiro, I., Bettencourt, S., and Sottomayor, M. (2016). Protoplast transformation as a plant-transferable transient expression system. *Methods Mol. Biol.* 1405, 137–148. doi: 10.1007/978-1-4939-3393-8_13

Dutta, I., Saha, P., and Das, S. (2008). Efficient agrobacterium-mediated genetic transformation of oilseed mustard [Brassica juncea (L.) czern.] using leaf piece explants. *In Vitro Cell. Dev. Biol. – Plant* 44, 401–411. doi: 10.1007/s11627-008-9150-1

Fahey, J. W., Wade, K. L., Stephenson, K. K., Panjwani, A. A., Liu, H., Cornblatt, G., et al. (2019). Bioavailability of sulforaphane following ingestion of glucoraphanin-rich broccoli sprout and seed extracts with active myrosinase: A pilot study of the effects of proton pump inhibitor administration. *Nutrients* 11(7), 1489. doi: 10.3390/nu11071489

Fu, Y. Y., Jia, S. R., and Lin, Y. (1985). Plant regeneration from mesophyll protoplast culture of cabbage (Brassica oleracea var 'capitata'). *Theor. Appl. Genet.* 71, 495–499. doi: 10.1007/bf00251195

Gasper, A. V., Al-Janobi, A., Smith, J. A., Bacon, J. R., Fortun, P., Atherton, C., et al. (2005). Glutathione s-transferase M1 polymorphism and metabolism of sulforaphane from standard and high-glucosinolate broccoli. *Am. J. Clin. Nutr.* 82, 1283–1291. doi: 10.1093/ajcn/82.6.1283

Georgikou, C., Buglioni, L., Bremerich, M., Roubicek, N., Yin, L., Gross, W., et al. (2020). Novel broccoli sulforaphane-based analogues inhibit the progression of pancreatic cancer without side effects. *Biomolecules* 10(5), 769. doi: 10.3390/biom10050769

Gould, J., Devey, M., Hasegawa, O., Ulian, E. C., Peterson, G., Smith, R. H., et al. (1991). Transformation of zea mays l. using agrobacterium tumefaciens and the shoot apex. *Plant Physiol.* 95, 426–434. doi: 10.1104/pp.95.2.426

Greb, T., and Lohmann, J. U. (2016). Plant stem cells. *Curr. Biol.* 26, R816–R821. doi: 10.1016/j.cub.2016.07.070

Han, F., Huang, J., Xie, Q., Liu, Y., Fang, Z., Yang, L., et al. (2021). Genetic mapping and candidate gene identification of BoGL5, a gene essential for cuticular wax biosynthesis in broccoli. *BMC Genomics* 22, 811. doi: 10.1186/s12864-021-08143-7

Huang, J., et al. (2021). Genetic diversity and population structure analysis of 161 broccoli cultivars based on SNP markers. *Hortic. Plant J.* 7, 423–433. doi: 10.1016/j.hpj.2021.05.002

Hu, Y., Song, D., Gao, L., Ajayo, B. S., Wang, Y., Huang, H., et al. (2020). Optimization of isolation and transfection conditions of maize endosperm protoplasts. *Plant Methods* 16, 96. doi: 10.1186/s13007-020-00636-y

Kang, L., Li, P., Wang, A., Ge, X., and Li, Z. (2017). A novel cytoplasmic Male sterility in brassica napus (inap CMS) with carpelloid stamens *via* protoplast fusion with Chinese woad. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00529

Keck, A. S., Qiao, Q., and Jeffery, E. H. (2003). Food matrix effects on bioactivity of broccoli-derived sulforaphane in liver and colon of F344 rats. *J. Agric. Food Chem.* 51, 3320–3327. doi: 10.1021/jf026189a

Kiełkowska, A., and Adamus, A. (2019). Peptide growth factor phytosulfokine-α stimulates cell divisions and enhances regeneration from b. oleracea var. capitata l. protoplast culture. *J. Plant Growth Regul.* 38, 931–944. doi: 10.1007/s00344-018-9903-y

Kong, W., Li, J., Qingyue, Y., Cang, W., Xu, R., Wang, Y., et al. (2016). Two novel flavin-containing monooxygenases involved in biosynthesis of aliphatic glucosinolates. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01292

Lee, Y. S., Ku, K. M., Becker, T. M., and Juvik, J. A. (2017). Chemopreventive glucosinolate accumulation in various broccoli and collard tissues: Microfluidic-based targeted transcriptomics for by-product valorization. *PloS One* 12, e0185112. doi: 10.1371/journal.pone.0185112

Li, Z., Mei, Y., Liu, Y., Fang, Z., Yang, L., Zhuang, M., et al. (2019). The evolution of genetic diversity of broccoli cultivars in China since 1980. *Sci. Hortic-Amsterdam* 250, 69–80. doi: 10.1016/j.scienta.2019.02.034

Li, Z., Zheng, S., Liu, Y., Fang, Z., Yang, L., Zhuang, M, et al. (2021). Characterization of glucosinolates in 80 broccoli genotypes and different organs using UHPLC-Triple-TOF-MS method. *Food Chem.* 334, 127519. doi: 10.1016/j.foodchem.2020.127519

Li, X., Sandgrind, S., Moss, O., Guan, R., Ivarson, E., Wang, E. S., et al. (2021). Efficient protoplast regeneration protocol and CRISPR/Cas9-mediated editing of glucosinolate transporter (GTR) genes in rapeseed (Brassica napus l.). *Front. Plant Sci.* 12, 680859. doi: 10.3389/fpls.2021.680859

Li, Z., Song, L., Liu, Y., Han, F., and Liu, W.. (2022). Effects of nanocarbon solution treatment on the nutrients and glucosinolate metabolism in broccoli. *Food Chem.* 15, 100429. doi: 10.1016/j.fochx.2022.100429

Liqing, Z., Bochu, W., Jing, Z., Lingxi, C., Chuanyun, D., Chuanren, D., et al. (2005). Protoplast isolation of callus in echinacea augustifolia. *Colloids Surf B Biointerfaces* 44, 1–5. doi: 10.1016/j.colsurfb.2005.05.002

Li, Z., Song, L., Liu, Y., Han, F., and Liu, W. (2022). Electrophysiological, morphologic, and transcriptomic profiling of the ogura-CMS, DGMS and maintainer broccoli lines. *Plants* 11. doi: 10.3390/plants11040561

Liu, Z., and Friesen, T. L. (2012). Polyethylene glycol (PEG)-mediated transformation in filamentous fungal pathogens. *Methods Mol. Biol.* 835, 365–375. doi: 10.1007/978-1-61779-501-5_21

Luo, G., Li, B., and Gao, C. (2022). Protoplast isolation and transfection in wheat. *Methods Mol. Biol.* 2464, 131–141. doi: 10.1007/978-1-0716-2164-6_10

Mackon, E., Ma, Y., Jeazet Dongho Epse Mackon, G. C., Li, Q., Zhou, Q, Liu, P., et al. (2021). Subcellular localization and vesicular structures of anthocyanin pigmentation by fluorescence imaging of black rice (Oryza sativa l.) stigma protoplast. *Plants (Basel)* 10(4), 685. doi: 10.3390/plants10040685

Mahn, A., and Castillo, A. (2021). Potential of sulforaphane as a natural immune system enhancer: A review. *Molecules* 26(3), 752. doi: 10.3390/molecules26030752

Marand, A. P., Chen, Z., Gallavotti, A., and Schmitz, R. J. (2021). A cis-regulatory atlas in maize at single-cell resolution. *Cell* 184, 3041–3055.e21. doi: 10.1016/j.cell.2021.04.014

Min, B. W., Cho, Y. N., Song, M. J., Noh, T. K., Kim, B. K., Chae, W. K., et al. (2007). Successful genetic transformation of Chinese cabbage using phosphomannose isomerase as a selection marker. *Plant Cell Rep.* 26, 337–344. doi: 10.1007/s00299-006-0247-x

Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. *Nat. Methods* 19, 534–546. doi: 10.1038/s41592-022-01409-2

Page, M. T., Parry, M. A. J., and Carmo-Silva, E. (2019). A high-throughput transient expression system for rice. *Plant Cell Environ.* 42, 2057–2064. doi: 10.1111/pce.13542

Radke, S. E., Turner, J. C., and Facciotti, D. (1992). Transformation and regeneration of brassica rapa using agrobacterium tumefaciens. *Plant Cell Rep.* 11, 499–505. doi: 10.1007/bf00236265

Reed, K. M., and Bargmann, B. O. R. (2021). Protoplast regeneration and its use in new plant breeding technologies. *Front. Genome Ed* 3. doi: 10.3389/fgeed.2021.734951

Rubel, M., Abuyusuf, M., Nath, U., Robin, A., Jung, H., Hoy, T., et al. (2020). Glucosinolate profile and glucosinolate biosynthesis and breakdown gene expression manifested by black rot disease infection in cabbage. *Plants* 9, 1–16. doi: 10.3390/plants9091121

Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* 179, 1444–1456. doi: 10.1104/pp.18.01482

Sahab, S., Hayden, M., Mason, J., and Spangenberg, G. (2019a). Mesophyll protoplasts and PEG-mediated transfections: Transient assays and generation of stable transgenic canola plants. *Methods Mol Biol* 1864, 131–152. doi: 10.1007/978-1-4939-8778-8_10

Sahab, S., Hayden, M. J., Mason, J., and Spangenberg, G. (2019b). Mesophyll protoplasts and PEG-mediated transfections: Transient assays and generation of stable transgenic canola plants. *Methods Mol. Biol.* 1864, 131–152. doi: 10.1007/978-1-4939-8778-8_10

Sivanandhan, G., Bae, S., Sung, C., Choi, S. R., Lee, G. J., Lim, Y. P., et al. (2021). Optimization of protoplast isolation from leaf mesophylls of Chinese cabbage (Brassica rapa ssp. pekinensis) and subsequent transfection with a binary vector. *Plants (Basel)* 10(12), 2636. doi: 10.3390/plants10122636

Sun, B., Yuan, Q., Zheng, H., Liang, S., Jiang, M., Wang, M. M., et al. (2019). An efficient and economical protocol for isolating, purifying and PEG-mediated transient gene expression of Chinese kale hypocotyl protoplasts. *Plants (Basel)* 8 (10), 385. doi: 10.3390/plants8100385

Tanveer, M., and Yousaf, U. (2020). *Chapter 23 - Plant single-cell biology and abiotic stress tolerance in Plant Life Under Changing Environment*. Ed. D. K. Tripathi, et al (Academic Press), 611–626.

Tao, H., Miao, H., Chen, L., Wang, M., Xia, C., Zeng, W., et al. (2022). WRKY33-mediated indolic glucosinolate metabolic pathway confers resistance against alternaria brassicicola in arabidopsis and brassica crops. *J. Integr. Plant Biol.* 64, 1007–1019. doi: 10.1111/jipb.13245

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

Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., et al. (2011). Glucosinolate biosynthetic genes in brassica rapa. *Gene* 487, 135–142. doi: 10.1016/j.gene.2011.07.021

Wang, J., Qiu, Y., Wang, X., Yue, Z., Yang, X., Chen, X., et al. (2017). Insights into the species-specific metabolic engineering of glucosinolates in radish (Raphanus sativus l.) based on comparative genomic analysis. *Sci. Rep-Uk* 7, 16040. doi: 10.1038/s41598-017-16306-4

Wang, Y., Zhang, Y.a., Dong, Y., Li, D., Shi, S., Li, S., et al. (2022). A highly efficient mesophyll protoplast isolation and PEG-mediated transient expression system in eggplant. *Scientia Hortic.* 304, 111303. doi: 10.1016/j.scienta.2022.111303

Wang, P., Pu, Y., Abid, M.A., Kang, L., Ye, Y., Zhang, M., et al. (2022). A rapid and efficient method for isolation and transformation of cotton callus protoplast. *Int. J. Mol. Sci.* 23(15), 8368. doi: 10.3390/ijms23158368

Wang, H., Wang, W., Zhan, J., Huang, W., and Xu, H. (2015). An efficient PEG-mediated transient gene expression system in grape protoplasts and its application in subcellular localization studies of flavonoids biosynthesis enzymes. *Scientia Hortic.* 191, 82–89. doi: 10.1016/j.scienta.2015.04.039

Xia, G. (2009). Progress of chromosome engineering mediated by asymmetric somatic hybridization. *J. Genet. Genomics* 36, 547–556. doi: 10.1016/s1673-8527(08)60146-0

Xie, Q., Zhao, Y., Liu, Y., Han, F., Liu, W., and Li, Z. (2022). Genetic diversity and DNA fingerprinting in broccoli carrying multiple clubroot resistance genes based on SSR markers. *Appl. Sci.* 12, 4754. doi: 10.3390/app12094754

Yang, Z.-N., Xu, Z.-H., and Wei, Z.-M. (1994). Cauliflower inflorescence protoplast culture and plant regeneration. *Plant Cell Tissue Organ Culture* 36, 191–195. doi: 10.1007/BF00037719

Yoo, S. D., Cho, Y. H., 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, Z., Duan, X., Luo, L., Dai, S., Ding, Z., and Xia, G. (2020). How plant hormones mediate salt stress responses. *Trends Plant Sci.* 25, 1117–1130. doi: 10.1016/j.tplants.2020.06.008

Zhang, Y., Su, J., Duan, S., Ao, Y., Dai, J., Liu, J., et al. (2011). A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant Methods* 7, 30. doi: 10.1186/1746-4811-7-30

Zhao, Y., Chen, Z., Chen, J., Chen, B., Tang, W., Chen, X., et al. (2021). Comparative transcriptomic analyses of glucosinolate metabolic genes during the formation of Chinese kale seeds. *BMC Plant Biol.* 21, 394. doi: 10.1186/s12870-021-03168-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 December 2022
DOI 10.3389/fpls.2022.1083528

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Tongbing Su,
Beijing Academy of Agricultural and Forestry Sciences, China
Xiaoming Song,
North China University of Science and Technology, China

*CORRESPONDENCE
Yun Zhang
zhangyun511@syau.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 29 October 2022
ACCEPTED 01 December 2022
PUBLISHED 19 December 2022

CITATION
Guan J, Li J, Yao Q, Liu Z, Feng H and Zhang Y (2022) Identification of two tandem genes associated with primary rosette branching in flowering Chinese cabbage.
Front. Plant Sci. 13:1083528.
doi: 10.3389/fpls.2022.1083528

Identification of two tandem genes associated with primary rosette branching in flowering Chinese cabbage

Jian Guan†, Jinyan Li†, Qingyu Yao, Zhiyong Liu, Hui Feng and Yun Zhang*

College of Horticulture, Shenyang Agricultural University, Shenyang, China

Branching is an important agronomic trait determining plant architecture and yield; however, the molecular mechanisms underlying branching in the stalk vegetable, flowering Chinese cabbage, remain unclear. The present study identified two tandem genes responsible for primary rosette branching in flowering Chinese cabbage by GradedPool-Seq (GPS) combined with Kompetitive Allele Specific PCR (KASP) genotyping. A 900 kb candidate region was mapped in the 28.0–28.9 Mb interval of chromosome A07 through whole-genome sequencing of three graded-pool samples from the F_2 population derived by crossing the branching and non-branching lines. KASP genotyping narrowed the candidate region to 24.6 kb. Two tandem genes, *BraA07g041560.3C* and *BraA07g041570.3C*, homologous to *AT1G78440* encoding GA2ox1 oxidase, were identified as the candidate genes. The *BraA07g041560.3C* sequence was identical between the branching and non-branching lines, but *BraA07g041570.3C* had a synonymous single nucleotide polymorphic (SNP) mutation in the first exon (290th bp, A to G). In addition, an ERE *cis*-regulatory element was absent in the promoter of *BraA07g041560.3C*, and an MYB *cis*-regulatory element in the promoter of *BraA07g041570.3C* in the branching line. Gibberellic acid (GA_3) treatment decreased the primary rosette branch number in the branching line, indicating the significant role of GA in regulating branching in flowering Chinese cabbage. These results provide valuable information for revealing the regulatory mechanisms of branching and contributing to the breeding programs of developing high-yielding species in flowering Chinese cabbage.

KEYWORDS

flowering Chinese cabbage, primary rosette branches, gene identification, tandem genes, GA2ox1 oxidase

Introduction

Brassica rapa is one of the most important *Brassica* species with a long cultivation history, has diverse and distinct morphological traits. Flowering Chinese cabbage [*Brassica rapa* L. ssp. *chinensis* (L.) Hanelt var. *parachinensis* (L.H. Bailey)] is that bolts readily (Wang and Kole et al., 2015). It is a stalk vegetable, and the stems with flower buds and leaves are consumed after cooking, especially in southern and central China and southeastern Asian countries. Typically, only one flowering stalk (main stem, without primary rosette branches) can be harvested from the common varieties, while several stalks are produced (12 stalks, primary rosette branches) on the local flowering Chinese cabbage variety named Zengcheng (Niu et al., 2019). Thus, the primary rosette branching trait influences plant architecture and yield in flowering Chinese cabbage. Therefore, identifying genes is important to develop varieties with multiple branches and meet the growing demand for flowering Chinese cabbage.

The activities of meristems, including apical, axillary and inflorescence meristems, basically determine the branching trait (Wang et al., 2018a). In flowering Chinese cabbage, the shoot apical meristem (SAM) turns into an inflorescence meristem that produces flowers directly or flower-bearing shoots after transitioning from the vegetative to the reproductive phase. The primary rosette branches grow from the axillary buds subtended by rosette leaves. Then, the secondary inflorescence branches grow at the axils of the cauline leaves on the elongated internodes of the main inflorescence stem, like that in *Arabidopsis thaliana* (Wang et al., 2018a; Fichtner et al., 2021). Studies in rice, *Arabidopsis*, and several other species characterizing the regulatory components of tiller or branch development have improved our understanding of branching (Wang and Li, 2011; Fichtner et al., 2021).

Typically, branching is a quantitative trait controlled by multiple genes and is susceptible to the environment (Ehrenreich et al., 2007; Kebrom et al., 2013) and plant hormones (Wang and Li, 2011). The hormonal control of bud outgrowth is complex and not yet fully understood. Auxin (Morris et al., 2005), cytokinin (Xu et al., 2015), strigolactone (De Jong et al., 2014), gibberellin (Martínez-Bello et al., 2015), abscisic acid (Holalu et al., 2020), and their interactions (Chen et al., 2013; Cao et al., 2017) have been reported to affect branching. Among these, gibberellin significantly influences the growth of branches and main stems. Okada et al. (2020) found that gibberellic acid (GA_3) application increased the number of lateral branches on apple trees. Meanwhile, the exogenous spraying of GA_3 rescued the dwarf phenotype of the legume's *msd1-2* (multi-*seeded1-2*) mutant (Li et al., 2021). Researchers have identified a few genes controlling branching in *Brassica juncea, Brassica napus*, non-heading Chinese cabbage, and purple flowering Chinese cabbage (Li, 2018; Muntha et al., 2019; Li et al., 2020a; Li et al., 2020b). However, branching in flowering Chinese cabbage has not been fully clarified.

Researchers recently proposed a new quantitative trait mapping technique called GradedPool-Seq (GPS) for rapidly mapping the quantitative trait loci (QTL), it can rapidly identify QTLs for complex traits comparing conditional methods (Wang et al., 2019). GPS with high-throughput sequencing scores and assigns the F_2 populations derived from a distant cross of parental lines exhibiting contrasting phenotypes into three or more graded groups based on phenotypic values. GPS has been successfully applied to dissect heterotic genes of thousand-grain weight, plant height, heading date, flag leaf angle, and tiller angle in rice. Moreover, the candidate intervals identified by the GPS method is consistent with the mapping interval obtained by the traditional method in rice (Wang et al., 2019).

Therefore, the present study used GPS with Kompetitive Allele Specific PCR (KASP) genotyping to map and identify the candidate genes associated with primary rosette branches in flowering Chinese cabbage. We further analyzed the similarities and variations in the full length and promoters of the candidate genes between the lines with different branching phenotypes. Finally, we investigated the role of GA_3 in regulating the development of multiple primary rosette branches in flowering Chinese cabbage. The findings of our study will provide novel insights into the mechanisms of branching and lay a foundation for developing flowering Chinese cabbage cultivars with multiple primary rosette branches.

Materials and methods

Plant materials and growing conditions

'CX020' (Branching line), a doubled haploid (DH) line derived by microspore culture from Zengcheng flowering Chinese cabbage [*Brassica rapa* L. ssp. *chinensis* (L.) Hanelt var. *parachinensis* (L.H. Bailey)] with multiple primary rosette branches at harvest, and 'CX010' (Non-branching line), a DH line derived from Guangdong flowering Chinese cabbage with one flower-bearing shoot and no primary rosette branch, were used as parents in this study. These DH parents were crossed to generate the F_1 and F_2 populations for the phenotypic and genetic analyses. These DH parents exhibited stable inheritance after multiple seasons of planting. All the plants ('CX010', 'CX020', and F_2 population) were grown at the Shenyang Agricultural University experimental base (Shenyang, China, *41°82'N, 123°24'E)* in 2019. The plants were sown on August 1st, and the number of primary rosette branches was analyzed on September 15th. The primary rosette branches were analyzed on the axillary branches subtended by rosette leaves 5 cm away from the cotyledonary node.

GradedPool-Seq

Three types of pools (50 plants each), including 'multiple primary rosette branching (12–15 branches)', 'less primary

rosette branching (1–3 branches)' and 'moderate primary rosette branching (7–8 branches)' plants were selected *via* phenotypic analysis from the F_2 population, consisting of 1050 individuals, were used for GPS analysis with the two parents. Young and fresh leaves of the parents and the selected F_2 individuals were harvested separately for total genomic DNA extraction using the Plant Genomic DNA Kit (Tiangen, Beijing, China), following the manufacturer's instructions. The quantity and quality of the DNA were ensured using spectrophotometric analysis and 2% agarose gel electrophoresis. The DNA samples were quantified using a Qubit fluorometer and pooled at equimolar concentrations to generate the 'multiple branching', 'less branching' and 'moderate branching' pools separately. Pair-end sequencing (PEN150) on an Illumina Novaseq system (Illumina, USA) was performed following the standard protocol. The sequencing data from each hybrid pool was merged and aligned to the reference genome (http://brassicadb.org/brad/datasets/pub/Genomes/Brassica_rapa/V3.0/) to calculate the depth of each variant. After filtering the variants with low quality and depth using the default parameters, Ridit analysis was performed to calculate the p-value for each variant. The sliding window size was set to 0.2 Mb, the threshold p-value to 10^{-8}, and the candidate regions to peak intervals to reduce background noise and identify the significant variants.

KASP genotyping

KASP was performed on a high-throughput Intelliqube genotyping platform for genotyping of the F_2 individuals. DNA was extracted from 150 F_2 individuals and the parents using the CTAB method. The DNA concentration and quality were assessed on a BioDropuLite microanalyzer (BioDrop, Britain), and the samples were diluted to a suitable concentration (5–10 ng/μL). KASP primers (Supplementary Table S1) were designed for 19 SNP loci (Supplementary Table S2) using Primer Premier 5.0 (Singh et al., 1998), and KASP assays were conducted in a 384-well plate format on a Hydrocycler (LGC, Middlesex, UK) using the following PCR protocol: 94 °C for 15 min; 94 °C for 20 s, 61 °C for 60 s (-1 °C/cycle, 10 cycles in total), and 94 °C for 20 s; 55 °C for 60 s (26 cycles). The components of the KASP reaction mixture are shown in Supplementary Table S3. The fluorescence signal generated was measured on an IntelliQube (LGC, Middlesex, UK). Finally, based on the KASP genotypic data and phenotypic data of the F_2 individuals, the QTL IciMapping software v4.2 (Liu et al., 2019) generated the linkage map to obtain the QTL and further narrow down the candidate interval.

Candidate gene analysis

Gene annotation information of the target region was obtained from the *Brassica* database (http://brassicadb.org/brad/index.php) and the *Arabidopsis* database (https://www.arabidopsis.org/). The candidate genes' full-length sequence and a 2000 bp long promoter sequence were amplified with the specific primers using PCR, and the amplicons were purified using a Gel Extraction Kit (CWBIO, Beijing, China). The purified products were introduced into the pGEM®-T Easy Vector (Promega, USA) and transformed into Top10 competent cells (CWBIO, Beijing, China). The colonies were sequenced at Sangon Biotech (Shanghai, China), and the sequences were aligned using DNAMAN 6.0 (Lynnon Biosoft, Canada). Meanwhile, PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict the *cis*-regulatory elements in the promoter regions of the candidate genes.

Expression analysis of the candidate genes

Total RNA was extracted from the whole roots, rosette stems, tender cauline leaves, flowers of top inflorescences and shoot tips of the parents. The rosette stem of CX020 at five different stages (every ten days), the first sample was taken when CX020 was in the fourth euphylla stage. The RNA was reverse transcribed using the FastKing RT Kit (Tiangen, Beijing). Real-time quantitative reverse transcription PCR (qRT-PCR) was carried out to determine the expression levels of the candidate genes using cDNA as the template with the SYBR Green PCR Master Mix on QuantStudio™ 6 Flex (Applied Biosystems, USA), maintaining three biological replicates per sample. The *Actin* gene was used as an internal reference control. The primers used for qRT-PCR were as follows: *BraA07g041560.3C* (F:5′-TGGAGATGATTACTGATGGGTTA-3′; R: 5′-ATTTTCGTGGATGAGAGGGC-3′); *BraA07g041570.3C* (F:5′-TCCTGGATTTCTGTCCCTTC-3′; R: 5′-ACCCTATGCTTCACGCTTTT-3′); *Actin* (F: 5′- ATCTACGAGGGTTATGCT-3′; R: 5′-CCACTGAGGACGATGTTT -3′). The relative gene expression levels were calculated following the $2^{-\Delta\Delta Ct}$ method.

Exogenous GA_3 and PAC treatment

We predicted that the genes affecting branching might be related to GA_3. An experiment was carried out by spraying GA_3 (750 mg/L) on the multiple primary rosette branching parent line CX020, and spraying Paclobutrazol (PAC, 0.3×10^{-3} mg/L on the non-primary rosette branching parent line CX010, using water as a control. The plants were sown on August 1st in green house at Shenyang Agricultural University under long-day conditions and GA_3 treatment was carried out on the fourth euphylla stage (August 15th); spraying was carried out every two days until phenotypes of fewer branches appeared. Twenty plants were maintained per treatment, using three biological

replicates. The primary rosette branches number was recorded when the branching phenotype is obvious.

Results

Phenotypic characterization of primary rosette branching in flowering Chinese cabbage

During the reproductive growth stage, 'CX010' (parent 1, Figure 1A) had only one main stalk but no primary rosette branches on the rosette stem, while 'CX020' (parent 2, Figure 1B) had 14 primary rosette branches that contributed to the yield. Moreover, significant differences were observed in the number of primary rosette branches between the non-branching line ('CX010') and the multiple branching line ('CX020'). Phenotypic segregation analysis showed that F_1 progeny had a moderate number of primary rosette branches. The number of primary rosette branches in the 1050 F_2 plants ranged from 0 to 15 and showed a normal distribution (Figure 1C). These observations indicated the role of QTLs in controlling the number of primary rosette branching in flowering Chinese cabbage.

Candidate region for primary rosette branching identified by GPS

The sequences obtained from the three pools (12–15 branches, 7–8 branches, and 1–3 branches) were mapped to the reference genome (http://brassicadb.org/brad/datasets/pub/Genomes/Brassica_rapaV3.0/) to estimate the allelic frequencies. Ridit analysis was implemented with allelic frequencies from three bulks to calculate p-values for each SNP. The background noise complicated the precise localization of the QTLs. Subsequently, the statistical noise-reducing strategy narrowed the interval to about 900 kb (28.0–28.9 Mb; significant peak) on chromosome A07, which was identified as the main QTL controlling the primary rosette branching. A few minor peaks also appeared on the other chromosomes that may be minor QTLs (Figure 2A). However, we further focused on the main QTL.

Further mapping of the primary rosette branching gene

A total of 178 genes were identified in the candidate region based on the gene information in the *Brassica* database (Supplementary Table S4). Then, to locate the candidate genes controlling the primary rosette branching in this region, we developed 18 KASP markers according to SNPs in this region. The genotype of 150 F_2 individuals and two parents was detected. Only one QTL was discovered associated with the number of primary rosette branches in the F_2 based on KASP, which accounted for 20.81% of the phenotypic variance. Markers A0716 and A0717, located at the two sides of the candidate gene, were the most closely associated with the candidate gene, and the physical distance between these markers was 24.6 kb, which contained five genes (Figure 2B).

Identification of the candidate genes related to primary rosette branching

Then, all the candidate genes in the candidate region were analyzed based on the *Brassica* database to identify the candidate genes that influence primary rosette branching. To further identify the candidate genes that influence primary rosette branching, five annotated genes in the candidate region were analyzed (Table 1). Among these, two tandem genes,

FIGURE 1
Phenotype of parent lines CX010, CX020 and distribution of the number of primary rosette branches in flowering Chinese cabbage. **(A)** Non-primary rosette branching line ('CX010'). **(B)** Multiple primary rosette branching line ('CX020') Bar = 10 cm. **(C)** Distribution of the number of primary rosette branches in 1050 individuals of F_2 population.

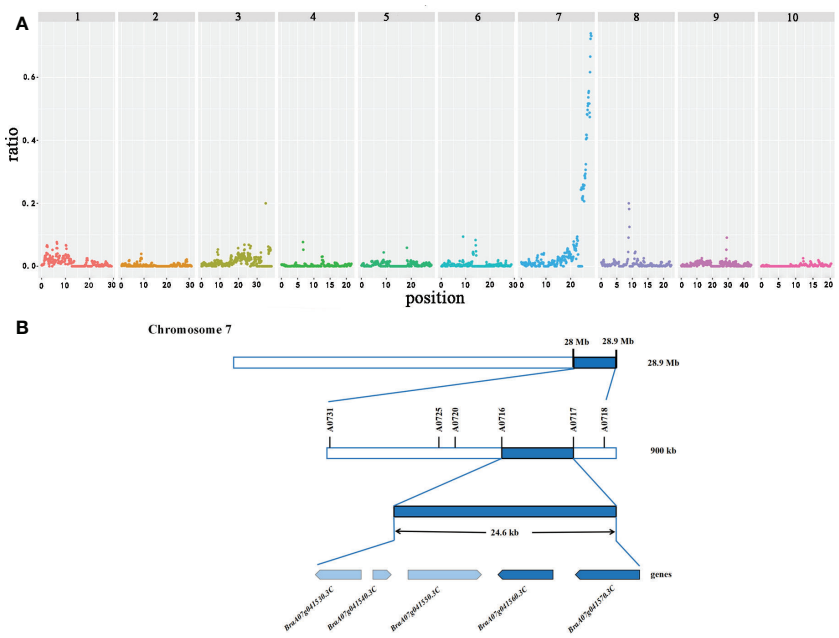


FIGURE 2
Candidate interval and genes identified by GPS on chromosome A07 in flowering Chinese cabbage. **(A)** The X-axis value is set at a midpoint at each defined genomic interval, and the Y-axis value corresponds to the ratio. **(B)** Candidate genes identified in the targeted interval.

BraA07g041560.3C and *BraA07g041570.3C*, were found homologous to the *Arabidopsis thaliana AT1G78440*, which encodes gibberellin 2-oxidase that acts on C19 gibberellins. In rice, *GA2oxs* influence the number of tillers (Lo et al., 2008). Therefore, we hypothesized that these two tandem genes might be associated with the number of primary rosette branches in flowering Chinese cabbage.

Candidate gene cloning and sequence analysis

These two full-length genes (*BraA07g041560.3C* and *BraA07g041570.3C*) were cloned to analyze the variances between the parents (Supplementary Figures S1, S2). Sequence analysis revealed that the full-length sequence of *BraA07g041560.3C* was identical in the parents, whereas *BraA07g041570.3C* had a synonymous SNP mutation from A to G at the 290th position (Supplementary Figure S2). The promoter sequence of *BraA07g041560.3C* had a 31 bp deletion in CX020 (Supplementary Figure S3). There were many differences of the promoter sequence in *BraA07g041560.3C* between CX020 and CX010 that resulted in the absence of an ERE *cis*-regulatory element and the position changes of many cis-acting elements in CX020 (Supplementary Figure S3; Figure 3A). Compared to CX010, CX020 had an SNP in the promoter of *BraA07g041570.3C*, 103 bp upstream of the translation initiation site, resulting in the absence of an MYB *cis*-regulatory element (Supplementary Figure S4; Figure 3B).

TABLE 1 Annotation of the genes within the mapped region on chromosome A07 in flowering Chinese cabbage.

Gene	Start	End	Gene annotations (BLASTX to *Arabidopsis thaliana*)	E value
BraA07g041530.3C	28005077	28006447	PGX2 is a cell wall protein that codes for a polygalacturonase.	0.0
BraA07g041540.3C	28013792	28014106	VQ motif-containing protein; (source: Araport11)	1.72^{e-51}
BraA07g041550.3C	28018304	28020467	Activates the latent peptidases DA1, DAR1 and DAR2 by mono-ubiquitination at multiple sites. Subsequently, these activated peptidases destabilize various positive regulators of growth.	5.64^{e-45}
BraA07g041560.3C	28023878	28025430	Encodes a gibberellin 2-oxidase that acts on C19 gibberellins.	0.0
BraA07g041570.3C	28036581	28038308	Encodes a gibberellin 2-oxidase that acts on C19 gibberellins.	0.0

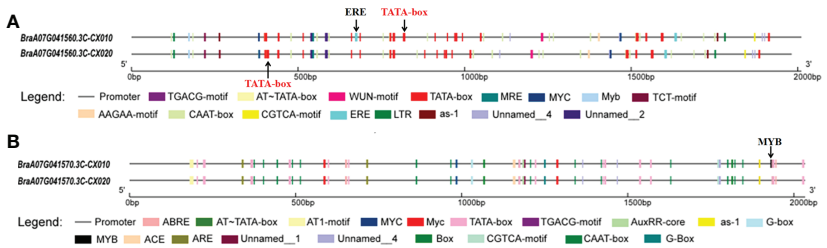


FIGURE 3
The *cis*-regulatory elements analysis of *BraA07g041560.3C* and *BraA07g041570.3C* promoters in parents CX010 and CX020. **(A)** Analysis of *BraA07g041560.3C* gene promoters. **(B)** Analysis of *BraA07g041570.3C* gene promoters.

Expression analysis of candidate genes

Further, qRT-PCR was used to detect the relative expression levels of *BraA07g041560.3C* and *BraA07g041570.3C* in CX010 and CX020. The relative expression level of *BraA07g041560.3C* in the stem and flower of CX020 was significantly higher than that in CX010, while that in the leaf and shoot tip of CX020 was markedly lower than that in CX010 (Figure 4A). The relative expression level of *BraA07g041570.3C* in CX020 was substantially higher in root, stem, flower, and shoot tip than that of CX010 but significantly lower in the leaf (Figure 4B). We further analyzed the differences in the expression levels of *BraA07g041560.3C* and *BraA07g041570.3C* in CX020 stem at different stages. The analysis revealed that the *BraA07g041560.3C* expression was the highest at the second stage (Figure 5A). Meanwhile, the *BraA07g041570.3C* expression level at the last stage was significantly different from those at the other stages (Figure 5B).

Phenotypic features after GA_3 and PAC treatment

GA_3 treatment significantly decreased the number of primary rosette branches (Figures 6A, B). On the contrary, after PAC treatment on the non-branching line CX010, the number of primary rosette branches were significantly increased (Figures 6C, D). The number of rosette branches significantly decreased from 10 to 0 after GA_3 treatment (Figure 7A), while the number of rosette branches significantly increased from 0 to 8 after PAC treatment (Figure 7B).

Discussion

Branching is an important trait that determines plant architecture, directly influences yield, and is closely related to

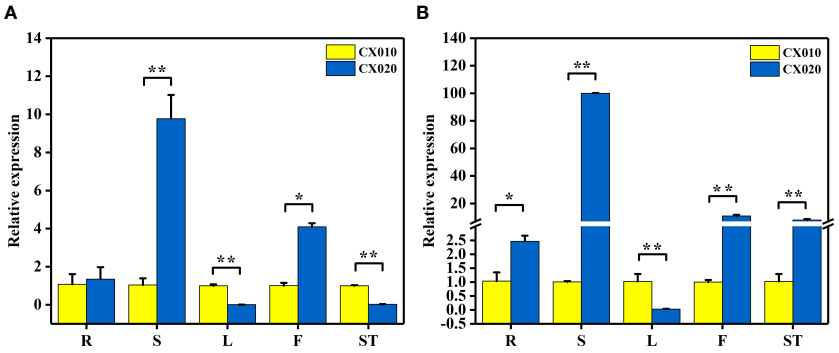


FIGURE 4
Expression levels of *BraA07g041560.3C* and *BraA07g041570.3C* in different tissues of CX010 and CX020 plants based on qRT-PCR. **(A)** The expression level of *BraA07g041560.3C*. **(B)** The expression level of *BraA07g041570.3C*. R, whole roots; S, rosette stems; L, tender cauline leaves; F, flowers of top inflorescences; ST, shoot tips. The data shown are the means of three replicates (± SD). * and ** indicate significant differences in expression levels at $P < 0.05$ and $P < 0.01$, respectively (Student's *t*-test).

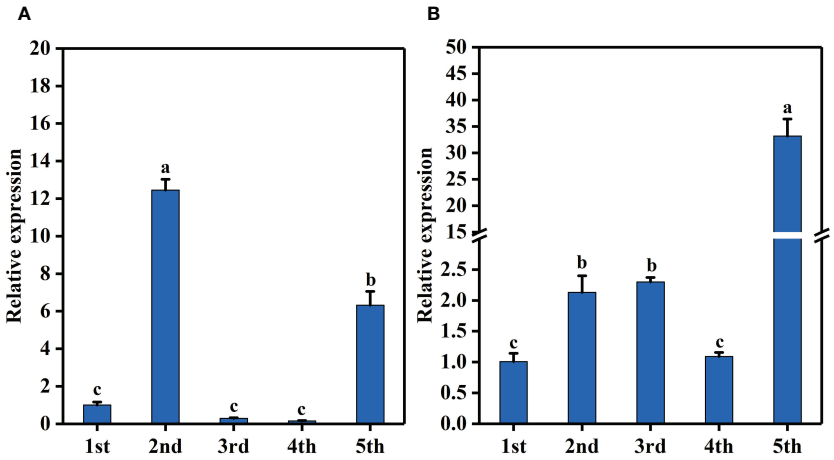


FIGURE 5
Expression levels of *BraA07g041560.3C* and *BraA07g041570.3C* in CX020 rosette stem at different stages based on qRT-PCR. **(A)** The expression level of *BraA07g041560.3C*. **(B)** The expression level of *BraA07g041570.3C*. The data shown are the means of three replicates (± SD). The different lowercase letters above the means are significantly different at P=0.05 level.

environmental adaptation (Teichmann and Muhr, 2015; Mathan et al., 2016). The primary rosette branches that contribute to yield in flowering Chinese cabbage are different from the rosette branches in non-heading Chinese cabbage and tillers in cereal crops. In non-heading Chinese cabbage, basal branches arise from axillary meristems in the leaf axils subtended by rosette leaves during the vegetative stage (Cao et al., 2016). Then the inflorescence branches grow out from the axillary meristems. In cereal crops, tillers arise from non-elongated internodes at the base of the parent shoot during the vegetative growth phase and survive even if the primary shoot dies because tillers generally produce adventitious roots (Kebrom et al., 2013). While, the branches in flowering Chinese cabbage develop only after transitioning to the reproductive phase, the stage at which tiller development ceases in cereal crops, such as wheat, barley, and rice. Studies have identified genes related to branching in *Brassica* crops. Li et al. predicted *BnaA09.ELP6* controls effective primary cauline branching (arising from the main stem) in *Brassica napus* (Li et al., 2020a). Meanwhile, shoot branching in non-heading Chinese cabbage (*Brassica rapa* ssp. *chinensis* Makino) is controlled by *BrSB9.1* (*Bra007056*), which is homologous to *MOC1* that controls tillering in rice (Li et al., 2020b). *Bra004212*, the homolog of *TCP1*, was identified as the candidate gene for tillering in purple flowering Chinese cabbage (Li, 2018). *PAT1* (Phytochrome A signal transduction 1), which belongs to the GRAS transcription factor family, negatively regulates branching in leafy *Brassica juncea* (Muntha et al., 2019). However, the genes controlling primary rosette branching in Chinese cabbage differ from these reported species. The present study the first time identified for two tandem genes, *BraA07g041560.3C* and *BraA07g041570.3C*, which are homologous to *AT1G78440* encoding *GA2ox1* oxidase, as potential candidate genes responsible for primary rosette branching.

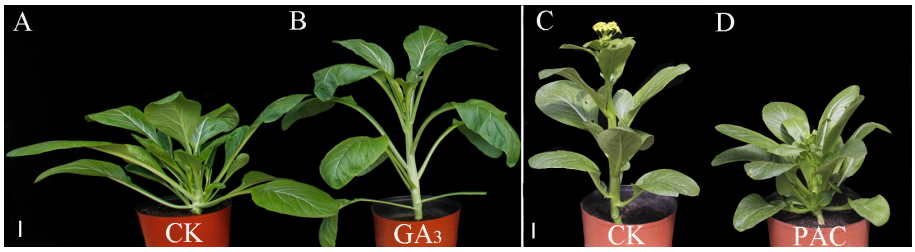


FIGURE 6
Phenotypes of CX020 and CX010 after treatments with exogenous GA_3 and PAC. **(A, B)** The CK and exogenous GA_3 treatment of CX020. Bar = 8 cm. **(C, D)** The CK and exogenous PAC treatment of CX010. Bar = 5 cm.

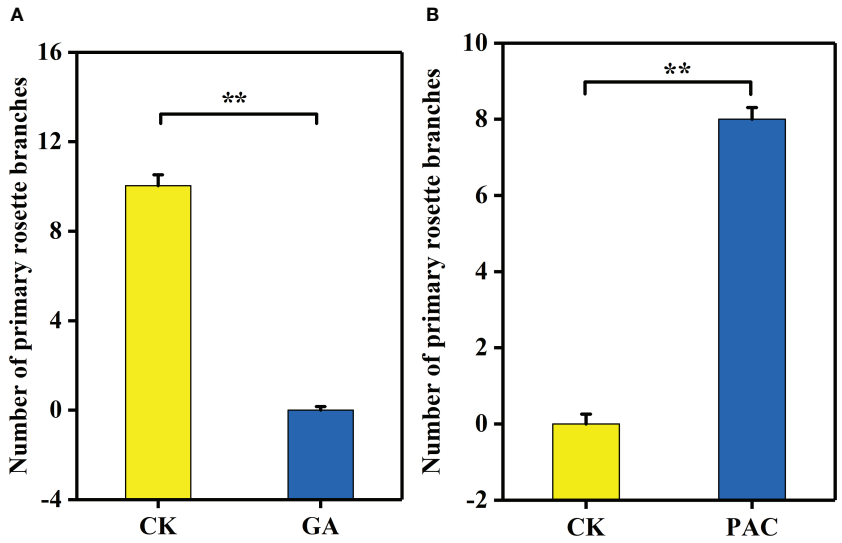


FIGURE 7
The number of primary rosette branches of CX020 and CX010 treated with exogenous GA_3 and PAC. **(A)** The CK and exogenous GA_3 treatment of CX020. **(B)** The CK and exogenous PAC treatment of CX010. The data shown are the means of three replicates (± SD). ** indicate significant differences in expression levels at $P < 0.01$ (Student's t-test).

Traditional QTL mapping using genetic map construction by traditional markers and phenotyping is a reliable approach to isolate genes or QTLs associated with agronomic traits (Li et al., 2020b). Hundreds of SSR (Simple Sequence Repeat), InDels, or SLAF (Specific-Locus Fragment)markers have been developed to construct a primary genetic linkage map based on the population of the F_2 generation, doubled haploid (DH) or recombinant inbred line (RIL) and locate QTL according to the phenotype (Liu et al., 2019). Usually, to narrow down the region and screen for a few candidate genes, a near-isogenic line (NIL) population is needed. However, the process of developing the NIL population is time-consuming and labor-intensive. GPS is a quick and efficient method to ascertain the genomic regions that harbor QTL for complex quantitative traits (Wang et al., 2019). It accelerates gene mapping by sequencing the graded pools; here, only the F_2 population is needed. Bulked-segregant analysis (BSA) also rapidly and effectively locates genes by constructing segregating F_2 populations from parents with significant phenotypic differences and selecting individuals with extreme traits to build a DNA pool for sequencing (Giovannoni et al., 1991; Michelmore et al., 1991; Takagi et al., 2013; Zou et al., 2016). However, the GPS has a higher resolution (~400 kb in rice) (Wang et al., 2019) than BSA (3 Mb) (Wang et al., 2018b; Liu et al., 2019; Yang et al., 2021). GPS identified a QTL controlling fruit size in melons, and traditional QTL mapping validated the results (Lian et al., 2021). Liu et al. (2020) successfully identified two candidate intervals controlling extremely late flowering in rice by GPS. In this work, one major candidate region associated with primary rosette branching was finally located in a 900 kb region on chromosome A07 based on three graded pools according to the primary rosette branching number of the F_2 population in flowering Chinese cabbage, with 178 genes. To our knowledge, this is the first report on the genetic control of primary rosette branching in flowering Chinese cabbage using GPS.

KASP genotyping based on uniplex SNP is a novel approach to fine map genes using F_2 individuals combined with their phenotype data (Xu et al., 2018; Cheng et al., 2021). Liu et al. (2019) narrowed the candidate region from 3.29 Mb to 790 kb by QTL analysis using KASP markers with 147 F_2 individuals in tomatoes and identified the *Cf-10* gene (*Cladosporium fulvum*). Lei et al. (2020) narrowed the genome interval from 4.17 Mb to 222 kb by 26 KASP markers genotyping with 199 individuals randomly selected from $F_{2:3}$ and identified a major QTL and candidate gene for salt tolerance in rice. In this study, we genotyped 150 F_2 individuals using 18 KASP markers and narrowed down the candidate region from 900 kb to 24.6 kb, with five genes.

GA2 oxidases are dioxygenases encoded by multiple genes and the key enzymes involved in gibberellin metabolism (Hedden and Phillips, 2000). GA2 oxidase typically transforms the bioactive GA1 and GA4 into the inactive catabolic metabolites GA8 and GA34, respectively, reducing the activity of GA and maintaining the balance between bioactive and inactive GA (Claeys et al., 2014). Studies have demonstrated the role of *GA2ox1* in regulating rosette branching in some crops. Higher expression levels of *GA2ox* genes have been correlated to low concentrations of bioactive GAs (Schomburg

et al., 2003; Dijkstra et al., 2008; Zhou et al., 2011). Meanwhile, turfgrass (*Paspalum notatum* Flugge) overexpressing *AtGA2ox1* had significantly lower levels of active GA but more tillers than wild-type plants (Agharkar et al., 2007). Overexpression of the *OsGA2ox* in rice increased tiller number (Lo et al., 2008), consistent with an increased number of tillers observed with the overexpression of *PvGA2ox5* and *PvGA2ox9* in switchgrass (*Panicum virgatum* L.) (Wuddineh et al., 2014). On the other hand, silencing of five *GA2ox* genes in tomatoes significantly increased GA4 content and inhibited lateral branches (Martínez-Bello et al., 2015). Therefore, we speculated that the two tandem genes, *BraA07g041560.3C* and *BraA07g041570.3C*, homologous to the *Arabidopsis thaliana* gene (*AT1G78440*) encoding a gibberellin 2-oxidase, found in the candidate region might be associated with primary rosette branching in flowering Chinese cabbage. While in the candicate region, the other three genes were not found to be related to branching. *BraA07g041530.3C* was the homologs gene of *AtPGX2*, which has been demonstrated regulating root hair development in response to phospho-starved (Zhang et al., 2022). *BraA07g041540.3C* was the homologs gene of *AtVQ10*, which has been shown to interact with *WRKY33* to affect plant sizes at mature stages in *Arabidopsis* (Cheng et al., 2012). *BraA07g041550.3C* was found to be homologous to *Arabidopsis At1g78420* (*DA2*), which encoded RING-type protein with E3 ubiquitin ligase activity and regulated seed size by restricting cell proliferation in the maternal integuments of developing seeds (Xia et al., 2013).

Typically, variations in the promoter regions may lead to changes in gene expression levels (Mito et al., 1996). The expression levels of these two genes were significantly higher in rosette stems of the branching line than in the non-branching line. Detailed analysis revealed differences in the promoter sequences between the parents. The promoter of the *BraA07g041560.3C* gene missed an ERE *cis*-regulatory element in branching line CX020, with a difference in the location of a TATA-box and the *cis*-regulatory element. However, there is no evidence for the role of the ERE *cis*-regulatory element in the formation of branching. Meanwhile, the promoter sequence of the *BraA07g041570.3C* gene lacked an MYB *cis*-regulatory element in CX020. Studies presented associated several genes belonging to the MYB family (MYB2, MYB37, and MYB181) with axillary meristem development and branching in *Arabidopsis* (Guo and Gan, 2011; Keller et al., 2006; Yang et al., 2018). AtMYB2 protein represses the formation of axillary meristems in response to salt and drought stresses (Jia et al., 2020). Meanwhile, the overexpression of *GmMYB181* in *Arabidopsis* altered the plant architecture, increased lateral branches, and reduced plant height (Yang et al., 2018). Therefore, we hypothesized that the *MYB* gene control primary rosette branching in flowering Chinese cabbage.

GAs are a large group of diterpenoid natural products characterized by tetracyclic 6-5-6-5 ring derived from ent-gibberellane (MacMillan and Takahashi, 1968; Peters, 2010). The biosynthesis of GA is a complex multi-step process requiring a variety of functional enzymes to catalyze the different intermediates (Wei et al., 2019). GA normally inhibits shoot branching, and plants overexpressing GA catabolic genes and GA-deficient mutants exhibit more shoot branching comparing to the wild-type (Silverstone et al., 1997; Agharkar et al., 2007; Lo et al., 2008). GA regulates internode elongation in rice, where the bioactive GA probably prevents from reaching the nodes below the shoot apex and inhibits internode elongation during the vegetative phase (Sakamoto et al., 2001).In this work, exogenous GA_3 application significantly reduced the rosette branches of flowering Chinese cabbage.

Conclusions

The present study identified two tandem genes, *BraA07g041560.3C* and *BraA07g041560.3C*, homologous to *AT1G78440* encoding GA2ox1 oxidase, as the candidates genes related to primary rosette branching in flowering Chinese cabbage. We detected differences in *cis*-regulatory elements in the promoter sequences of *BraA07g041560.3C* and *BraA07g041570.3C* between the branching and non-branching lines, which indicated the role of genes in regulating branching. These results provide valuable information for revealing the branching regulatory mechanisms in flowering Chinese cabbage. Further studies should investigate and conform the possible function and the promoter activity of *BraA07g041560.3C* and *BraA07g041570.3C* in flowering Chinese cabbage.

Data availability statement

The data presented in the study are deposited in the SRA repository of the National Center for Biotechnology Information, accession number PRJNA908111 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA908111.

Ethics statement

The authors note that this research was performed and reported in accordance with ethical standards of scientific conduct.

Author contributions

YZ and HF designed the experiments. JG and JL conducted the experiments and wrote the manuscript. JG, JL, and QY performed the data analysis. YZ revised the manuscript. All authors reviewed and approved this manuscript.

Funding

The research was supported by the National Natural Science Foundation of China (Grant No. 31972404) and Science Study Foundation of Liaoning (LJKZ0640) and China Agriculture research system (CARS-23).

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.2022.1083528/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The full-length of gene *BraA07g041560.3C*.

SUPPLEMENTARY FIGURE 2
The full-length of gene *BraA07g041570.3C*.

SUPPLEMENTARY FIGURE 3
The promoter sequence of *BraA07g041560.3C*.

SUPPLEMENTARY FIGURE 4
The promoter sequence of *BraA07g041570.3C*.

SUPPLEMENTARY TABLE 1
KASP primer sequence used in this study.

SUPPLEMENTARY TABLE 2
SNP site information.

SUPPLEMENTARY TABLE 3
The components of the KASP reaction mixture.

SUPPLEMENTARY TABLE 4
The 178 genes identified in the candidate region based on the gene information in the *Brassica* database.

References

Agharkar, M., Lomba, P., Altpeter, F., Zhang, H. N., Kenworthy, K., and Lange, T. (2007). Stable expression of *AtGA2ox1* in a low-input turfgrass (*Paspalum notatum* flugge) reduces bioactive gibberellin levels and improves turf quality under field conditions. *Plant Biotechnol. J.* 5, 791–801. doi: 10.1111/j.1467-7652.2007.00284.x

Cao, X. W., Cui, H. M., Yao, Y., Xiong, A. S., Hou, X. L., and Li, Y. (2017). Effects of endogenous hormones on variation of shoot branching in a variety of non-heading Chinese cabbage and related gene expression. *J. Plant Biol.* 60, 343–351. doi: 10.1007/s12374-016-0124-2

Cao, X. W., Cui, H. M., and Li, J. (2016). Heritability and gene effects for tiller number and leaf number in non-heading Chinese cabbage using joint segregation analysis. *Scientia Horticulturae* 203, 199–206. doi: 10.1016/j.scienta.2016.03.018

Cheng, Z. K., Liu, Z. G., Xu, Y. C., Ma, L. L., Chen, J. Y., Gou, J. Q., et al. (2021). Fine mapping and identification of the candidate gene *BFS* for fruit shape in wax gourd (*Benincasa hispida*). *Theor. Appl. Genet.* 134, 3983–3995. doi: 10.1007/s00122-021-03942-8

Cheng, Y., Zhou, Y., and Yang, Y. (2012). Structural and functional analysis of VQ motif-containing proteins in arabidopsis as interacting proteins of WRKY transcription factors. *Plant Physiol. (Rockville)* 159 (2), 810. doi: 10.1104/pp.112.196816

Chen, X. L., Zhou, X. Y., Xi, L., Li, J. X., Zhao, R. Y., Ma, N., et al. (2013). Roles of *DgBRC1* in regulation of lateral branching in chrysanthemum (*Dendranthema* x *grandiflora* cv. jinba). *PLoS One* 8, e61717. doi: 10.1371/journal.pone.0061717

Claeys, H., De Bodt, S., and Inzé, D. (2014). Gibberellins and DELLAs: central nodes in growth regulatory networks. *Trends Plant Sci.* 19, 231–239. doi: 10.1016/j.tplants.2013.10.001

De Jong, M., George, G., Ongaro, V., Williamson, L., Willetts, B., Ljung, K., et al. (2014). Auxin and strigolactone signaling are required for modulation of Arabidopsis shoot branching by nitrogen supply. *Plant Physiol.* 166, 384–395. doi: 10.1104/pp.114.242388

Dijkstra, C., Adams, E., Bhattacharya, A., Page, A. F., Anthony, P., Kouurmptli, S., et al. (2008). Over-expression of a gibberellin 2-oxidase gene from *Phaseolus coccineus* L. enhances gibberellin inactivation and induces dwarfism in *Solanum* species. *Plant Cell Rep.* 27, 463–470. doi: 10.1007/s00299-007-0471-z

Ehrenreich, I. M., Stafford, P. A., and Purugganan, M. D. (2007). The genetic architecture of shoot branching in *Arabidopsis thaliana*: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping. *Genetics* 176, 1223–1236. doi: 10.1534/genetics.107.071928

Fichtner, F., Barbier, F. F., Annunziata, M. G., Feil, R., Olas, J. J., Mueller-Roeber, B., et al. (2021). Regulation of shoot branching in *Arabidopsis* by trehalose 6-phosphate. *New Phytol.* 229, 2135–2151. doi: 10.1111/nph.17006

Giovannoni, J. J., Wing, R. A., Ganal, M. W., and Tanksley, S. D. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 19, 6553–6558. doi: 10.1093/nar/19.23.6553

Guo, Y. F., and Gan, S. S. (2011). *AtMYB2* regulates whole plant senescence by inhibiting cytokinin-mediated branching at late stages of development in *Arabidopsis*. *Plant Physiol.* 156, 1612–1619. doi: 10.1104/pp.111.177022

Hedden, P., and Phillips, A. L. (2000). Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.* 5, 523–530. doi: 10.1016/S1360-1385(00)01790-8

Holalu, S. V., Reddy, S. K., Blackman, B. K., and Finlayson, S. A. (2020). Phytochrome interacting factors 4 and 5 regulate axillary branching *via* bud abscisic acid and stem auxin signalling. *Plant,Cell Environ.* 43, 2224–2238. doi: 10.1111/pce.13824

Jia, T., Zhang, K., Li, F., Huang, Y. F., Fan, M. M., Huang, T., et al. (2020). The AtMYB2 inhibits the formation of axillary meristem in Arabidopsis by repressing RAX1 gene under environmental stresses. *Plant Cell Rep* 39, 1755–1765. doi: 10.1007/s00299-020-02602-3

Kebrom, T. H., Spielmeyer, W., and Finnegan, E. J. (2013). Grasses provide new insights into regulation of shoot branching. *Trends Plant Sci.* 18, 41–48. doi: 10.1016/j.tplants.2012.07.001

Keller, T., Abbott, J., Moritz, T., and Doerner, P. (2006). *Arabidopsis REGULATOR OF AXILLARY MERISTEMS1* controls a leaf axil stem cell niche

and modulates vegetative development. *Plant Cell* 18, 598–611. doi: 10.1105/tpc.105.038588

Lei, L., Zheng, H. L., Bi, Y. L., Yang, L. M., Liu, H. L., Wang, J. G., et al. (2020). Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* l.) using QTL-seq and RNA-seq. *Rice* 13, 22. doi: 10.1186/s12284-020-00416-1

Li, Y. X. (2018). *Genetic analysis and QTL mapping of tillering in purple caitai* (Wuhan(HB: Huazhong Agricultural University).

Lian, Q., Fu, Q. S., Xu, Y. Y., Hu, Z. C., Zheng, J., Zhang, A. A., et al. (2021). QTLs and candidate genes analyses for fruit size under domestication and differentiation in melon (*Cucumis melo* l.) based on high resolution maps. *BMC Plant Biol.* 21, 126. doi: 10.1186/s12870-021-02904-y

Li, B., Gao, J., Chen, J., Wang, Z., Shen, W., Yi, B., et al. (2020a). Identification and fine mapping of a major locus controlling branching in *Brassica napus*. *Theor. Appl. Genet.* 133, 771–783. doi: 10.1007/s00122-019-03506-x

Li, W., Ma, Q., Yin, P., Wen, J., Pei, Y., Niu, L., et al. (2021). The GA 20-oxidase encoding gene *MSD1* controls the main stem elongation in medicago truncatula. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.709625

Li, P., Su, T., Zhang, B., Li, P., Xin, X., Yue, X., et al. (2020b). Identification and fine mapping of *qSB.A09*, a major QTL that controls shoot branching in brassica rapa ssp. *chinensis* makino. *Theor. Appl. Genet.* 133, 1055–1068. doi: 10.1007/s00122-020-03531-1

Liu, J., Gong, J. Y., Wei, X., Yang, S. H., Huang, X. H., Li, C., et al. (2020). Dominance complementation of *Hd1* and *Ghd8* contributes to extremely late flowering in two rice hybrids. *Mol. Breed.* 40, 76. doi: 10.1007/s11032-020-01162-4

Liu, G., Zhao, T., You, X. Q., Jiang, J. B., Li, J. F., and Xu, X. Y. (2019). Molecular mapping of the *Cf-10* gene by combining SNP/InDel-index and linkage analysis in tomato (*Solanum lycopersicum*). *BMC Plant Biol.* 19, 15. doi: 10.1186/s12870-018-1616-7

Lo, S. F., Yang, S. Y., Chen, K. T., Hsing, Y. I., Zeevaart, J. A., Chen, L. J., et al. (2008). A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell* 20, 2603–2618. doi: 10.1105/tpc.108.060913

MacMillan, J., and Takahashi, N. (1968). Proposed procedure for the allocation of trivial names to the gibberellins. *Nature* 217, 170–171. doi: 10.1038/217170a0

Martínez-Bello, L., Moritz, T., and López-Díaz, I. (2015). Silencing C_{19}-GA 2-oxidases induces parthenocarpic development and inhibits lateral branching in tomato plants. *J. Exp. Bot.* 66, 5897–5910. doi: 10.1093/jxb/erv300

Mathan, J., Bhattacharya, J., and Ranjan, A. (2016). Enhancing crop yield by optimizing plant developmental features. *Development* 143, 3283–3294. doi: 10.1242/dev.134072

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828

Mito, N., Wimmers, L., and Bennett, A. (1996). Sugar regulates mRNA abundance of h+-ATPase gene family members in tomato. *Plant Physiol.* 112, 1229–1236. doi: 10.1104/pp.112.3.1229

Morris, S. E., Cox, M. C., Ross, J. J., Krisantini, S., and Beveridge, C. A. (2005). Auxin dynamics after decapitation are not correlated with the initial growth of axillary buds. *Plant Physiol.* 138, 1665–1672. doi: 10.1104/pp.104.058743

Muntha, S. T., Zhang, L., Zhou, Y., Zhao, X., Hu, Z., Yang, J., et al. (2019). Phytochrome a signal transduction 1 and CONSTANS-LIKE 13 coordinately orchestrate shoot branching and flowering in leafy *Brassica juncea*. *Plant Biotechnol. J.* 17, 1333–1343. doi: 10.1111/pbi.13057

Niu, L. J., Shi, F. Y., Feng, H., and Zhang, Y. (2019). Efficient doubled haploid production in microspore culture of zengcheng flowering Chinese cabbage (*Brassica campestris* L. ssp. *Chinensis* [L.] makino var. *utilis* tsen et Lee). *Scientia Hortic.* 245, 57–64. doi: 10.1016/j.scienta.2018.09.076

Okada, K., Wada, M., Takebayashi, Y., Kojima, M., Sakakibara, H., Nakayasu, M., et al. (2020). Columnar growth phenotype in apple results from gibberellin deficiency by ectopic expression of a dioxygenase gene. *Tree Physiol.* 40, 1205–1216. doi: 10.1093/treephys/tpaa049

Peters, R. J. (2010). Two rings in them all: The labdane-related diterpenoids. *Natural Product Rep.* 27, 1521–1530. doi: 10.1039/C0NP00019A

Sakamoto, T., Kobayashi, M., Itoh, H., Tagiri, A., Kayano, T., Tanaka, H., et al. (2001). Expression of a gibberellin 2-oxidase gene around the shoot apex is related to phase transition in rice. *Plant Physiol.* 125, 1508–1516. doi: 10.1104/pp.125.3.1508

Schomburg, F. M., Bizzell, C. M., Lee, D. J., Zeevaart, J. A., and Amasino, R. M. (2003). Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15, 151–163. doi: 10.1105/tpc.005975

Silverstone, A. L., Mak PiuYing, A., Casamitjana Martinez, E., and Sun, T. (1997). The new RGA locus encodes a negative regulator of gibberellin response in *Arabidopsis thaliana*. *Genetics* 146 (3), 1087–1099. doi: 10.1038/s41598-017-10823-y

Singh, V. K., Mangalam, A. K., Dwivedi, S., and Naik, S. (1998). Primer premier: program for design of degenerate primers from a protein sequence. *Biotechniques* 24, 318–319. doi: 10.2144/98242pf02

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

Teichmann, T., and Muhr, M. (2015). Shaping plant architecture. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00233

Wang, X. W., and Kole, C. (2015). *The Brassica rapa Genome: Economic/Academic Importance of Brassica rapa* (New York: Springer-Verlag Berlin Heidelberg), 1–2. doi: 10.1007/978-3-662-47901-8

Wang, Y., and Li, J. Y. (2011). Branching in rice. *Curr. Opin. Plant Biol.* 14, 94–99. doi: 10.1016/j.pbi.2010.11.002

Wang, N., Liu, Z. Y., Zhang, Y., Li, C. Y., and Feng, H. (2018b). Identification and fine mapping of a stay-green gene (*Brnye1*) in pakchoi (*Brassica campestris* L. ssp. *chinensis*). *Theor. Appl. Genet.* 131, 673–684. doi: 10.1007/s00122-017-3028-8

Wang, B., Smith, S. M., and Li, J. (2018a). Genetic regulation of shoot architecture. *Annu. Rev. Plant Biol.* 69, 437–468. doi: 10.1146/annurev-arplant-042817-040422

Wang, C., Tang, S., Zhan, Q., Hou, Q., Zhao, Y., Zhao, Q., et al. (2019). Dissecting a heterotic gene through GradedPool-seq mapping informs a rice-improvement strategy. *Nat. Communication* 10 (1), 2982. doi: 10.1038/s41467-019-11017-y

Wei, C. H., Zhu, C. Y., Yang, L. P., Zhao, W., Ma, R. X., Li, H., et al. (2019). A point mutation resulting in a 13 bp deletion in the coding sequence of *Cldf* leads to a GA-deficient dwarf phenotype in watermelon. *Horticulture Res.* 6, 132. doi: 10.1038/s41438-019-0213-8

Wuddineh, W., Mazarei, M., Zhang, J. Y., Poovaiah, C., Mann, D., Ziebell, A., et al. (2014). Identification and overexpression of *gibberellin 2-oxidase* (*GA2ox*) in switchgrass (*Panicum virgatum* l.) for improved plant architecture and reduced biomass recalcitrance. *Plant Biotechnol. J.* 13, 636–647. doi: 10.1111/pbi.12287

Xia, T., Li, Na., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., et al. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in arabidopsis. *Plant Cell* 25, 3347–3359. doi: 10.1105/tpc.113.115063

Xu, X., Ji, J., Xu, Q., Qi, X., Weng, Y., and Chen, X. (2018). The major-effect quantitative trait locus *CsARN6.1* encodes an AAA ATPase domain-containing protein that is associated with waterlogging stress tolerance by promoting adventitious root formation. *Plant J.* 93, 917–930. doi: 10.1111/tpj.13819

Xu, J., Zha, M., Li, Y., Ding, Y., Chen, L., Ding, C., et al. (2015). The interaction between nitrogen availability and auxin, cytokinin, and strigolactone in the control of shoot branching in rice (*Oryza sativa* l.). *Plant Cell Rep.* 34, 1647–1662. doi: 10.1007/s00299-015-1815-8

Yang, S. J., Tian, X. X., Wang, Z. Y., Wei, X. C., Zhao, Y. Y., Su, H. N., 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. doi: 10.3389/fpls.2021.646222

Yang, H., Xue, Q., Zhang, Z. Z., Du, J. Y., Yu, D. Y., and Huang, F. (2018). GmMYB181, a soybean R2R3-MYB protein, increases branch number in transgenic *Arabidopsis*. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01027

Zhang, Q., Deng, A. W., and Xiang, M. (2022). The root hair development of pectin polygalacturonase PGX2 activation tagging line in response to phosphate deficiency. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.862171

Zhou, B., Peng, D., Lin, J., Huang, X., Peng, W., He, R., et al. (2011). Heterologous expression of a gibberellin 2-oxidase gene from *Arabidopsis thaliana* enhanced the photo-synthesis capacity in *Brassica napus* l. *J. Plant Biol.* 54, 23–32. doi: 10.1007/s12374-010-9139-2

Zou, C., Wang, P., and Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* 14, 1941–1955. doi: 10.1111/pbi.12559

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fpls.2022.1067508

OPEN ACCESS

EDITED BY
Xiaodong Yang,
Yangzhou University, China

REVIEWED BY
Dengfeng Hong,
Huazhong Agricultural University, China
Zhansheng Li,
Insititute of Vegetables and Flowers (CAAS), China
Kunjiang Yu,
Guizhou University, China

*CORRESPONDENCE
Dong-Hwan Kim
dhkim92@cau.ac.kr

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 11 October 2022
ACCEPTED 29 December 2022
PUBLISHED 18 January 2023

CITATION
Kim JA, Moon H, Kim HS, Choi D, Kim N-S, Jang J, Lee SW, Baskoro Dwi Nugroho A and Kim D-H (2023) Transcriptome and QTL mapping analyses of major QTL genes controlling glucosinolate contents in vegetable- and oilseed-type *Brassica rapa* plants.
Front. Plant Sci. 13:1067508.
doi: 10.3389/fpls.2022.1067508

Transcriptome and QTL mapping analyses of major QTL genes controlling glucosinolate contents in vegetable- and oilseed-type *Brassica rapa* plants

Jin A. Kim[1†], Heewon Moon[2†], Hyang Suk Kim[1], Dasom Choi[2], Nan-Sun Kim[1], Juna Jang[1], Sang Woo Lee[2], Adji Baskoro Dwi Nugroho[2] and Dong-Hwan Kim[2*]

[1]Department of Agricultural Biotechnology, National Institute of Agricultural Science, Rural Development Administration, Jeonju, Jeollabuk-do, Republic of Korea, [2]Department of Plant Science and Technology, Chung-Ang University, Anseong, Republic of Korea

Glucosinolates (GSLs) are secondary metabolites providing defense against pathogens and herbivores in plants, and anti-carcinogenic activity against human cancer cells. Profiles of GSLs vary greatly among members of genus *Brassica*. In this study, we found that a reference line of Chinese cabbage (*B. rapa* ssp. *pekinensis*), 'Chiifu' contains significantly lower amounts of total GSLs than the oilseed-type *B. rapa* (*B. rapa* ssp. *trilocularis*) line 'LP08'. This study aimed to identify the key regulators of the high accumulation of GSLs in *Brassica rapa* plants using transcriptomic and linkage mapping approaches. Comparative transcriptome analysis showed that, in total, 8,276 and 9,878 genes were differentially expressed between 'Chiifu' and 'LP08' under light and dark conditions, respectively. Among 162 *B. rapa* GSL pathway genes, 79 were related to GSL metabolism under light conditions. We also performed QTL analysis using a single nucleotide polymorphism-based linkage map constructed using 151 F_5 individuals derived from a cross between the 'Chiifu' and 'LP08' inbred lines. Two major QTL peaks were successfully identified on chromosome 3 using high-performance liquid chromatography to obtain GSL profiles from 97 F_5 recombinant inbred lines. The MYB-domain transcription factor gene *BrMYB28.1* (Bra012961) was found in the highest QTL peak region. The second highest peak was located near the 2-oxoacid-dependent dioxygenase gene *BrGSL-OH.1* (Bra022920). This study identified major genes responsible for differing profiles of GSLs between 'Chiifu' and 'LP08'. Thus, our study provides molecular insights into differences in GSL profiles between vegetative- and oilseed-type *B. rapa* plants.

KEYWORDS

glucosinolate, *Brassica rapa*, QTL mapping, *BrMYB28.1*, transcriptome

1 Introduction

Secondary metabolites of plants have diverse functions throughout the plant's lifespan. A lack of plant secondary metabolites does not lead to immediate death, but can affect the survival and reproduction of plant species over the long term (Isah, 2019). Many secondary metabolites play important roles in plant defense systems against a variety of environmental stresses including salt, drought, heat, wounding, and attacks from pathogens (Dixon, 2001). Furthermore, plant secondary metabolites determine important aspects of human food quality, such as taste and flavor (Verpoorte and Memelink, 2002).

Glucosinolates (GSLs), a type of plant secondary metabolite, are mainly produced in crop plants of the Brassicaceae family, and help the plants resist stresses including attack by insects and herbivores (Halkier and Gershenzon, 2006; Hopkins et al., 2009). More than 130 GSLs have been identified in the Brassicaceae family (Agerbirk and Olsen, 2012; Thinh Nguyen et al., 2020). GSLs are not only important molecules for plant defense, but are also reported to have anti-cancer, anti-inflammatory, and other health benefits in humans (Verhoeven et al., 1996; Keck and Finley, 2004). Some GSLs, such as glucoraphanin (GRA), glucoalyssin (GAS), gluconapin (GNP), neoglucobrassicin (NGB), and gluconasturtiin (GNT), are beneficial, whereas hydrolysis products from progoitrin (PRO), epiprogoitrin (epiPRO), and gluconapoleiferin (GNL) can cause goiter in animals (Sønderby et al., 2010b). GSLs are derived from amino acids and can be divided into aliphatic (derived from Met, Leu, Ala, Ile, and Val), indolic (derived from Trp), and benzenic (derived from Phe and Tyr) GSLs based on their amino acid precursors (Grubb and Abel, 2006).

Structural differences among GSL compounds are driven mainly by variations of the genes involved in the initial side chain elongation and secondary modification stages (Halkier and Gershenzon, 2006). GSL biosynthetic processes have been intensively studied in the model plant *Arabidopsis thaliana* (Sønderby et al., 2010a). A small subgroup of R2R3-type myeloblastosis (MYB) transcription factors (TFs) plays an important role in the regulation of GSL metabolism. For example, *Arabidopsis* has three *MYB* TF genes (*MYB28*, *MYB29*, and *MYB76*) controlling aliphatic GSL biosynthesis, and another three (*MYB34*, *MYB51*, and *MYB122*) regulating indolic GSL biosynthesis (Gigolashvili et al., 2007; Hirai et al., 2007; Sønderby et al., 2007). Double mutants of the *MYB28* and *MYB29* genes, designated *myb28; myb29*, exhibited severe reduction of aliphatic GSLs (Li et al., 2013). Another TF, the DNA-binding with one finger (DOF) domain-containing TF *OBP2* (also referred to as *AtDof1.1*, AT1G07640), positively regulated aliphatic GSL biosynthesis (Skirycz et al., 2006). All of these TFs function as positive regulators of GSL biosynthesis.

In addition to these TFs, many genes encoding catalytic enzymes are involved in the aliphatic and indolic GSL biosynthetic pathways. GSL biosynthesis generally consists of four stages: the 'chain elongation' stage of amino acid precursors such as methionine (Met) and phenylalanine (Phe); the 'core structure formation' stage; the 'secondary modification' stage; and the 'breakdown' stage (Wittstock and Halkier, 2002). Initial stage 'side-chain elongation' starts from the side chain elongation of precursor amino acids, which involves the enzymes METHYLTHIOALKYL MALATE SYNTHASE 1 (MAM1), MAM2, MAM3, BRANCED-CHAIN AMINOTRANSFERASE 3 (BCAT3), BRANCED-CHAIN AMINOTRANSFERASE 4 (BCAT4), BILE ACID TRANSPORTER 5 (BAT5), IPMDH1 (ISOPROPYLMALATE DEHYDROGENASE 1), IPMI1 (ISOPROPYLMALATE ISOMERASE 1) and IPMI2. GSLs with various types of side-chain can be produced depending on allelic variations and/or even different developmental tissues. After the 'side chain elongation' stage, CYTOCHROME P450 enzymes, including CYP79F1 and CYP79F2 (aliphatic GSL pathway) and CYP79B2 and CYP79B3 (indolic GSL pathway) convert elongated amino acids into aldoxime which is further converted into aci-nitro compounds by CYP83A1 (aliphatic GSLs) and CYP83B1 (indolic GSLs). Subsequently, the aci-nitro compound is further modified through a series of catalytic activities by the enzymes SUR1, UGT74B1, UGT74C1, and ST5a/b/c (SOT16/18/17). After 'core structure formation', desulfo-GSLs undergo the 'secondary modification' process which involves FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASES (FMO GS-OXs), ALKENYL HYDROXALKYL PRODUCING 2 (AOP2), AOP3, and GLUCOSINOLATE HYDROXYLASE (GSL-OH) for aliphatic GSLs (Zhang et al., 2015; Kakizaki et al., 2017). The 'secondary modification' of indolic GSLs is processed by a small group of cytochrome P40 monooxygenase family proteins, including CYTOCHROME P450 MONOOXYGENASE 81 SUBUNIT 1 (CYP81F1) to CYP81F4, and INDOLE GLUCOSINOLATE O-METHYLTRANSFERASE 1 (IGMT1) and IGMT2. Later, GSL compounds are further hydrolyzed by plant enzymes, commonly called myrosinases including THIOGLUCOSIDE GLUCOHYDROLASE 1 (TGG1) to TGG5, EPITHIOSPECIFIER (ESP), and NITRILE SPECIFIER PROTEIN (NSP1) to NSP5 (aliphatic GSLs) and PENETRATION 2 (PEN2), PEN3, and CADMIUM SENSITIVE 1 (CAD1) (indolic GSLs).

It was previously reported that level of GSL compounds and the expression of GSL biosynthetic genes are positively affected by light (Schuster et al., 2006; Huseby et al., 2013). For instance, GSLs are highly synthesized during day time and significantly reduced in night time in a daily basis. Expression of MYB TF genes and their downstream GSL biosynthetic genes were also influenced by the presence of light. For example, expression of *MYB* TF genes (*MYB28*, *MYB29*, and *MYB76*) and GSL biosynthetic genes (*CYP79F1*, *SOT17*, *SOT18* etc.) involved in the aliphatic GSL biosynthesis were significantly reduced in the absence of light (Huseby et al., 2013). It indicated that light signaling positively stimulate expression of GSL pathway genes. Furthermore, light-stimulated increase of GSLs is also in an agreement with the fact

Abbreviations: Glucosinolates (GSLs), Glucoraphanin (GRA), Glucoalyssin (GAS), Gluconapoleiferin (GNA), Gluconapin (GNP), Glucobrassicanapin (GBN), Progoitrin (PGT), Glucoerucin (GER), Glucobrassicin (GBC), 4-Hydroxyglucobrassicin (4-HGB), 4-Methoxyglucobrassicin (4-MTGB), Neoglucobrassicin (NGB), Gluconasturtiin (GNT), isothiocyanates (ITC), Ultra-high performance liquid chromatography (UHPLC), Quantitative reverse transcription PCR (RT-qPCR), Transcription factors (TFs), Flavin-monooxygenase glucosinolate S-oxygenases (FMO GS-OXs), Alkenylhydroxalkyl producing (AOP), Glucosinolate hydroxylase (GSL-OH), Cytochrome P450 monooxygenase 81 subunit (CYP81F), Indole glucosinolate O-methytransferase (IGMT), Thioglucoside glucohydrolase (TGG), Epithiospecifier (ESP), Nitrile specifier protein (NSP), Penetration 2 (PEN2), Penetration 3 (PEN3), and Cadmium sensitive 1 (CAD1).

that glucose and sulfur, two essential precursors for GSL biosynthesis are highly synthesized during day time by photosynthesis and sulfate assimilation, respectively (Koprivova and Kopriva, 2014).

During last decades, many studies investigated the GSL profiles of various *Brassica* crops, such as *Brassica rapa*, Chinese kale, broccoli, cabbage, and cauliflower (Vallejo et al., 2002; Xu et al., 2006; Bellostas et al., 2007; Padilla et al., 2007; Jia et al., 2009; Sun et al., 2011; Wang et al., 2012). For example, total 16 GSLs were identified in *B. rapa* from seeds and leaves of adult stage plant, which displayed two aliphatic GSLs, gluconapin (GNP) and glucobrassicanapin (GBN) most abundant. Indolic and aromatic GSLs levels were low and showed few differences among different *B. rapa* varieties (Padilla et al., 2007). Bellostas et al. (2007) have investigated the GSLs profiles of five varieties of *B. oleracea* (white cabbage, red cabbage, Savoy cabbage, broccoli, and cauliflower) in sprouting stage. The concentration of alkyl-GSLs in these *B. oleracea* cultivars decreased, whereas glucobrassicin (GBS) significantly increased throughout the sprouting period. In broccoli, GSLs were evaluated in several cultivars (Vallejo et al., 2002). Dominant GSLs in all broccoli cultivars were glucoraphanin (GRA) and glucobrassicin (GBS) (Vallejo et al., 2003; Schonhof et al., 2004; Wang et al., 2012). GSLs profiles of Chinese Kale (*Brassica oleracea* var. *acephala*) is only limitedly informed. Sun et al. (2011) investigated the GSLs in three edible parts of Chinese kale (sprout, tender rosette leaf, and bolting stem). Thirteen GSLs, including eight aliphatic GSLs, four indole GSLs, and one aromatic GSL were identified in Chinese kale. The aliphatic GSLs were the most abundant GSLs in Chinese kale, with gluconapin (GNP) being the most abundant individual GSLs.

Brassica rapa has many subspecies with marked morphological variations (Seo et al., 2017), including oilseed crop yellow sarson (*B.rapa* ssp. *trilocularis*) for the production of seed oil and leafy vegetable type Chinese cabbage (*B.rapa* ssp. *pekinensis*) for leaf consumption. It has great differences between both types in terms of phenotypic traits. One of the Chinese cabbage vegetable type, 'Chiifu' line is self-incompatible and flowers late, requiring vernalization for bolting, and has a wide leaf shape. Meanwhile, the yellow sarson, 'LP08' line is self-compatible and flowers rapidly, and has narrow leaves with serrated margins.

In *B. rapa*, recent study quantified GSL levels of the eight subspecies of *B. rapa* which exhibited variable GSL levels ranging from 4.42 $\mu mol \cdot g^{-1}$ dw of pak choi (*B. rapa* ssp. *chinensis*) to 53.51 $\mu mol \cdot g^{-1}$ dw of yellow sarson (*B. rapa* ssp. *trilocularis*) (Seo et al., 2017). GSL profiling using the doubled haploid (DH) lines produced between high GSLs [*B. rapa* ssp *trilocularis* (yellow sarson)] and low GSLs [*B. rapa* ssp. *chinensis* (pak choi)] parents identified GSL-specific recombinant block on A03 (12.9 Mb ~ 23.2 Mb) chromosome based on SNP and InDels (Soundararajan et al., 2021). Similar to previous reports (Seo et al., 2017; Soundararajan et al., 2021), we also found that total amounts of GSLs were significantly higher in the oilseed-type yellow sarson 'LP08' (*B. rapa* ssp. *trilocularis*) line than the vegetable-type Chinese cabbage 'Chiifu' (*B. rapa* ssp. *pekinensis*) line. In this study, to elucidate the major gene or genes responsible for the difference in GSL contents between these two lines, we employed two comprehensive approaches: comparative transcriptome and quantitative trait locus (QTL) mapping. This study reveals the molecular mechanisms underlying the differing GSL contents between these two inbred lines, i.e., the molecular mechanisms of GSL pathways in *B. rapa* plants.

2 Materials and methods

2.1 Plant materials and growth conditions

For transcriptome analysis, the Chinese cabbage (B. rapa ssp. pekinensis) 'Chiifu' inbred line and yellow sarson (B. rapa ssp. trilocularis) 'LP08' line were used in this study. For the mapping population, 151 individuals in the F5 generation of a *Brassica rapa* segregating population were obtained by crossing the 'Chiifu' and 'LP08' lines in the greenhouse of Rural Development Administration of Korea as described previously (Kim et al., 2017; Kim et al., 2022). This population was used for QTL mapping of GSL contents.

Considering that GSL content is highly affected by a diversity of environmental stresses, in this study, we wanted to exclude environmental stress factors as much as possible and identify genetic factor(s) contributing to different GSL amounts between Chiifu and LP08. Thus, we grew seedling plants in a plant culture dish containing MS media in an environment-controlled growth chamber (22°C, 16h light/8h dark photoperiod). For quantification of GSL contents, 97 F_5 seedlings of recombinant inbred lines (RILs), as well as the parental lines 'Chiifu' and 'LP08', were used for extraction of GSLs. Seeds were surface-sterilized, spread on half-strength Murashige and Skoog agar medium and stored in the dark at 4°C for 3 days for stratification. Seedlings were grown for 1 week in a growth chamber at 22°C under long-day (LD) (16-h light/8-h dark) conditions. At least 5 seedling plants per line were harvested for high-performance liquid chromatography (HPLC) analysis.

2.2 Construction of genotyping-by-sequencing (GBS) libraries

All 151 F_5 lines, as well as the two parental lines, were subjected to GBS using the Illumina NextSeq500 sequencing platform, as described previously (Kim et al., 2022). GBS libraries were sequenced on the NextSeq500 system (Illumina, USA) and single-end reads 150 bp in length were obtained. After sequencing of GBS libraries, the raw reads were de-multiplexed to sort samples using the GBSX tool (Herten et al., 2015). The *Brassica rapa* reference genome (Brassica_rapa.Brapa_1.0.dna.toplevel.fa) was obtained from the EnsemblPlants genome database (https://plants.ensembl.org/index.html). After de-multiplexing, single-end reads were mapped to the *B. rapa* reference genome using Bowtie2 (Langmead and Salzberg, 2012). For calling of single nucleotide polymorphism (SNP) variants, the Genome Analysis Toolkit (GATK) and Picard tools (McKenna et al., 2010) packages were used. Local realignment of reads was performed to correct misalignment resulting from the presence of indels, using the GATK 'RealignerTargetCreator' and 'Indoleamine' sequence data processing tools. Lastly, the GATK 'HaplotypeCaller' and 'SelectVariants' tools were used for SNP variant calling.

2.3 Linkage map construction and QTL mapping

The variant call format SNP data were transformed into the input format using customized code for the R/Qtl package (eGenome, Republic of Korea). Markers with a duplicated pattern or distorted segregation ratio, as estimated using the Chi-square test with Bonferroni correction, were removed. The est.map function of R/Qtl was used to construct the linkage map, with the Kosambi map function converting genetic distances into recombination fractions (p-value threshold of 1e-06, and EM iterations in 1000 times). Subsequently, the composite interval mapping function of the R/Qtl package was used for QTL mapping with the Kosambi function. Regions inferred from peak positions with local maximum limit of detection (LOD) values exceeding the threshold determined using 1,000 permutation tests were considered significant QTLs. To detect QTLs responsible for the differing GSL profiles of the two parental lines, the GSL contents of 97 F_5 lines and the parental lines were measured and applied to the SNP linkage map. The LOD score threshold value for significance (α = 0.05) was estimated based on 1,000 permutation tests. Peaks exceeding the estimated LOD threshold value were selected for further analysis.

2.4 Extraction of GSLs and HPLC

Plants were grown at 22°C under a 16-h light/8-h dark photoperiod. One-week-old plants were harvested at ZT4 (4 h after light on) and then immediately ground in liquid nitrogen. GSLs were extracted as desulfo-glucosinolates (DS-GSLs), as reported previously (Han et al., 2019). Approximately 500 mg of fresh sample was incubated with 70% methanol at 70°C for 25 min to deactivate myrosinases. Contents of DS-GSLs were analyzed through ultra-high-performance liquid chromatography (UHPLC; 3000 U-HPLC system; ThermoFisher Scientific, USA). As a standard, sinigrin (0.5mg/ml) injection was used in all analyses (Sigma-Aldrich, USA). The DS-GSLs were resolved with a C18 reverse phase column (Zorbax XDB-C18, 4.6 × 250 mm, 5-μm particle size; Agilent, USA) with a water and acetonitrile gradient system. Samples were injected and maintained at a flow rate of 0.5 ml/min. Peaks were identified using standard compounds (Phytoplan, Germany). The samples were analyzed independently (three replicates) and the results are presented in nmol/g based on fresh weight (FW).

2.5 Analysis of GSLs through liquid chromatography coupled to diode array detection and electrospray ionization mass spectrometry

DS-GSLs were analyzed using the Accela UHPLC system (ThermoFisher Scientific) fitted with an ion trap mass spectrometer (LTQ Velos Pro; ThermoFisher Scientific). The samples were resolved using a C18 reverse phase column (Zorbax XDB-C18, 4.6 × 250 mm, 5-μm particle size; Agilent) with water and acetonitrile as the mobile phase, and measured in negative ion mode ([M-H]-). Mass spectrometry was conducted with the following settings: capillary temperature, 275°C; capillary voltage, 5kV; source heater temperature, 250°C; sheath gas flow, 35 arb; auxiliary gas flow, 5 arb; and spectral scanning range, m/z 100–1,500.

2.6 RNA sequencing and library construction

Total RNA was extracted from 1-week-old seedlings grown at 22°C under light (16-h light/8-h dark photoperiod) or dark conditions. Three biological replicates were harvested at each time point and frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, Germany) and subsequently treated with DNase I (NEB, USA) to remove contaminating genomic DNA. Purified total RNA was used for construction of RNA-seq libraries using the TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Paired-end sequencing was performed on the HiSeq 2500 system (Illumina).

2.7 Sequence alignment and analysis

Prior to alignment of RNA-seq reads to the *Brassica rapa* reference genome, FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was employed to evaluate the quality of the RNA-seq reads. Reads with > 90% Q values > 30 were filtered and used only for genome alignment. The *B. rapa* FASTA genome file (Brassica_rapa.Brapa_1.0.dna.toplevel.fa) and gff3 file (Brassica_rapa.Brapa_1.0.54.gff3) were downloaded from the EnsemblPlants genome database. TopHat2 mapping software was employed with the default parameters for alignment of reads to the reference genome (Kim et al., 2013). Digital read counts were obtained using featureCounts (Liao et al., 2014) and subsequently analyzed for differentially expressed genes (DEGs) using edgeR. The cut-off for DEGs was a two-fold difference ($p < 0.05$). A multi-dimensional scaling (MDS) plot, correlation map, and PlotSmear results were produced using R software (ver. 3.6.1; https://www.rstudio.com/products/rpackages/) packages. The web-based tool VENNY was used to generate a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). The web-based tool ShinyGO (ver. 0.61) was used for Gene Ontology (GO) enrichment analysis (http://bioinformatics.sdstate.edu/go/). Hierarchical clustering heatmap analysis was performed using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) and the JAVA TreeView program (Saldanha, 2004). Mapping results were visualized with the Integrative Genomics Viewer (IGV) program of the Broad Institute (Thorvaldsdóttir et al., 2013).

2.8 RNA extraction and quantitative reverse transcription polymerase chain reaction analysis

Total RNA was extracted from 1-week-old seedlings grown at 22°C under light (16-h light/8-h dark photoperiod) or dark conditions using the RNeasy Plant Mini Kit (Qiagen). Extracted RNA was treated with DNase I (NEB) to remove contaminating genomic DNA.

Approximately 5 μg of total RNA was used for cDNA synthesis with EasyScript reverse transcriptase (TransGen Biotech, China). qRT-PCR was performed using Solg 2× Real-Time PCR Smart Mix (SolGent, Republic of Korea) on a LineGene 9600 Plus Real-Time PCR system (BIOER, China), according to the manufacturer's instructions. qRT-PCR was conducted under the following conditions: denaturation at 95°C for 12 min, followed by 50 cycles of amplification (95°C for 15 s, 60°C for 25 s, 72°C for 35 s) and sampling at 72°C. The relative transcript level of each gene was determined through comparison with that of *BrPP2Aa* (Bra012474), a housekeeping gene consistently expressed in our RNA-seq analysis. The primers were designed based on sequences obtained from the *B. rapa* genome database (BRAD; http://brassicadb.cn). The primers used for qPCR analysis are presented in Supplementary Table S6. Student's *t*-test was used for statistical analysis (*P < 0.05, **P < 0.01, ***P < 0.001).

2.9 Sequencing of the *BrMYB28.1* and *BrGSL-OH.1* genes

PCR amplification was performed to clone two major QTL candidate genes, *BrMYB28.1* and *BrGSL-OH.1*, using gene-specific primers (Supplementary Table S6). PCR reaction was conducted with the following conditions: denaturation at 94°C for 5 min, followed by three initial cycles of amplification (94°C for 30 s, 55°C for 35 s, 72°C for 5 min), 33 additional cycles of amplification (94°C for 30 s, 64°C for 35 s, 72°C for 4 min 30 s) and a final extension step at 72°C for 10 min. The PCR product was extracted from the agarose gel after 1% agarose gel electrophoresis and then purified with the Dyne Power Gel Extraction Kit (Dynebio, Republic of Korea). The purified PCR products were inserted into the pPZP211 plant expression vector using the In-fusion HD cloning kit (Takara Bio, Japan). Entire genomic sequences of *BrMYB28.1* and *BrGSL-OH.1* were obtained through Sanger sequencing (Bionics, Republic of Korea) using the series of primers listed in Supplementary Table S6.

3 Results

3.1 Comparison of levels and compositions of GSLs between 'Chiifu' and 'LP08'

The vegetable-type 'Chiifu' and oilseed-type 'LP08' inbred lines grown under LD conditions were harvested to quantify GSL compounds using HPLC. Twelve GSL compounds representing all three GSL groups (aliphatic, indolic, and aromatic) were successfully identified using our detection system (Supplementary Figure S1). Among these 12 GSLs, 7 aliphatic, 4 indolic, and 1 aromatic GSL compound were identified (Supplementary Figure S1 and Table S1). The aromatic GSL compound GNT was detected at a very low level, and was therefore neglected in further analyses. Total GSL levels were apparently higher in 'LP08' than 'Chiifu' (Figure 1A and Supplementary Table S1). For example, the total GSL levels were 5,434.3 nmol/g FW in 'LP08' and 2,869.9 nmol/g FW in 'Chiifu'; thus, 'LP08' had around double the GSL level of 'Chiifu' (Figure 1B and Supplementary Table S1). In the LD sample, 94.74% of GSLs in 'LP08' and 91.37% in 'Chiifu' were aliphatic, while the remaining GSLs were indolic (Figure 1B). While the total amount of aliphatic GSLs in 'LP08' was almost double that in 'Chiifu', the total amounts of indolic GSLs did not differ dramatically between the two inbred lines, although the difference was found to be statistically significant (Figure 1A and Supplementary Table S1). Taken together, these results indicate that, in *B. rapa* seedlings, the majority of total GSLs are aliphatic GSLs.

Among aliphatic GSLs, two compounds (GNP and GER: glucoerucin) were present at significantly higher levels in 'LP08' than 'Chiifu' (Figure 1C). In particular, in 'LP08', GNP accounted for an overwhelming proportion of aliphatic GSLs (97.5% of total aliphatic GSLs) (Figure 1D). These results indicate that the higher level of total GSLs in 'LP08' is attributable to high abundance of GNP. Meanwhile, among aliphatic GSLs in 'Chiifu', five compounds (PGT, GRA, GAS, GNA: gluconapoleiferin, and GBN: glucobrassicanapin) had higher levels in 'Chiifu' than 'LP08' (Figure 1C). Among these compounds, GRA accounted for 50.0% (1,310.5 nmol/g FW) of total aliphatic GSLs, while GNP and GBN accounted for 30.3% (793.5 nmol/g FW) and 13.2% (347.2 nmol/g FW) of total GSLs, respectively (Figure 1D).

Two indolic GSLs (4-methoxyglucobrassicin: 4-MTGB and NGB) had higher levels in 'LP08' than 'Chiifu', while two other GSL compounds (glucobrassicin [GBC] and 4-hydroxyglucobrassicin [4-HGB]) were present at higher levels in 'Chiifu' than 'LP08', demonstrating a dynamic compositional difference of indolic GSLs between the two lines (Figure 1E). In terms of total indolic GSL levels, 'Chiifu' and 'LP08' had similar amounts (241.0 and 280.5 nmol/g FW, respectively; Figure 1F and Supplementary Table S1). However, detailed analysis of the composition of indolic GSL compounds revealed differing profiles (Figure 1F). For example, in 'Chiifu' 4-HGB was dominant (49.2%) among the four measured indolic GSL compounds. Meanwhile, NGB was dominant (55.7%) among the four indolic GSLs in 'LP08'. This indicates that 'Chiifu' and 'LP08' have differing compositional profiles for both aliphatic and indolic GSLs. Taken together, these results show that 'LP08' had a higher abundance of total GSLs than 'Chiifu', which can be attributed to greater accumulation of GNP, whereas several aliphatic GSL compounds were present in large proportions in 'Chiifu'.

3.2 RNA-seq and GO analyses

To identify the candidate genes responsible for the observed differences in amounts and profiles of GSLs between 'Chiifu' and 'LP08', RNA-seq was performed using 'Chiifu' and 'LP08' grown under light and dark conditions. MDS and correlation heatmap analyses of RNA-seq samples showed close clustering within each sample group, indicating that RNA-seq libraries were properly generated for 'Chiifu' and 'LP08' (Supplementary Figures S2A, B). We isolated DEGs between the two parental lines based on comparison of pairwise samples (Supplementary Figure S2C). Under light conditions, 3,054 and 5,222 genes were up- and down-

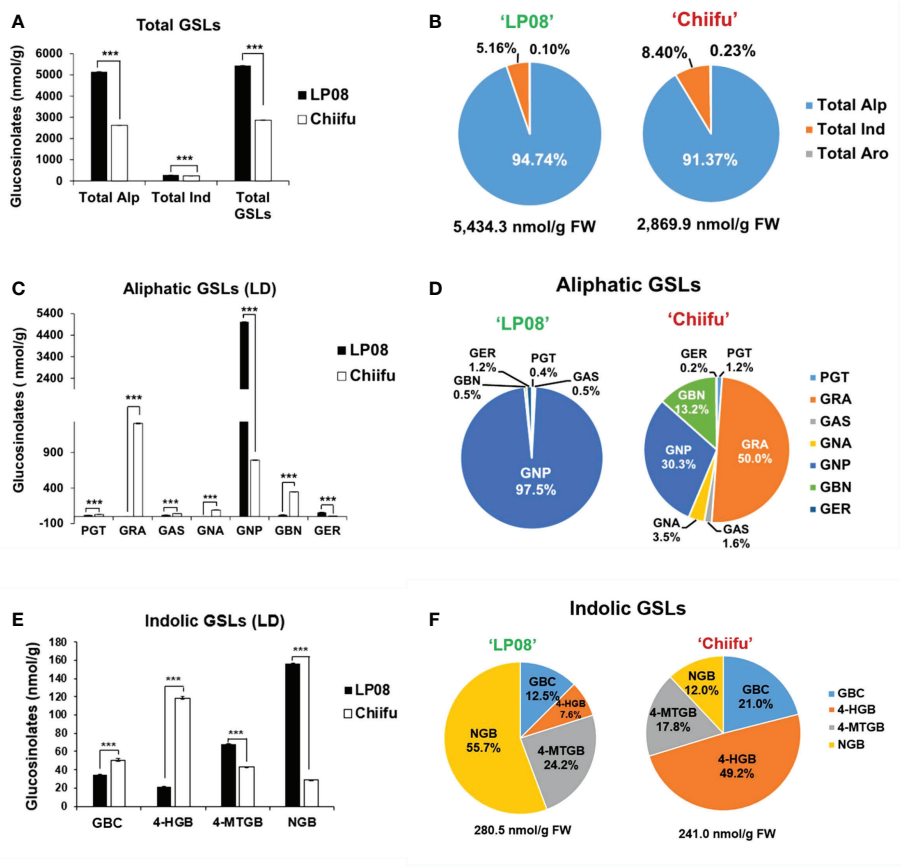


FIGURE 1
Measurement of GSL compounds in 'Chiifu' and 'LP08'. **(A)** Comparison of total GSL amounts (aliphatic, indolic, and aromatic GSL compounds) in 'Chiifu' and 'LP08' plants grown under long-day **(LD)** growth conditions. **(B)** Pie charts showing aliphatic (Alp), indolic (Ind), and aromatic (Aro) GSL compounds as percentages of total GSL content. Aliphatic GSL compounds constituted the majority of the GSLs in both the 'Chiifu' and 'LP08' lines. **(C)** Bar graph showing the amounts of seven aliphatic GSL compounds in 'Chiifu' and 'LP08'. Student's *t*-test was used to calculate statistically significant differences (***$p \leq 0.001$). **(D)** Pie charts showing the amounts of individual aliphatic GSLs in 'Chiifu' and 'LP08'. PGT, progoitrin; GRA, glucoraphanin; GAS, glucoalyssin; GNA, gluconapoleiferin; GNP, gluconapin; GBN, glucobrassicanapin; and GER, glucoerucin **(E)** Bar graph showing the amounts of four indolic GSL compounds in 'Chiifu' and 'LP08'. **(F)** Pie charts showing the amounts of four indolic GSL compounds in 'Chiifu' and 'LP08'. GBC, glucobrassicin; 4-HGB, 4-hydroxyglucobrassicin; 4-MTGB, 4-methoxyglucobrassicin; NGB, neoglucobrassicin.

regulated, respectively, in 'LP08' compared to 'Chiifu'. Under dark conditions, 4,659 and 5,219 genes were up- and down-regulated, respectively, in 'LP08' relative to 'Chiifu' (Figure 2A and Supplementary Table S3).

GO analysis allows for functional annotation by classifying individual genes based on their biological, cellular, and molecular functions. Thus, lists of up- and down-regulated genes in 'LP08' compared to 'Chiifu' grown under light and dark conditions were assessed with the ShinyGO analysis tool. We set 30% as an arbitrary threshold for significance. Interestingly, GO analysis of up-regulated genes in 'LP08' grown under light conditions revealed four categories related to GSL metabolism among the top 10 enriched categories (Supplementary Figure S3A). This result indicates that GSL-related metabolism is affected significantly more in 'LP08' compared to 'Chiifu'. We also obtained the top 10 categories of down-regulated genes in 'LP08' compared to 'Chiifu'. However, we detected no significant enrichment of GO categories above the threshold (Supplementary Figure S3B). These results suggest that GSL metabolism is an actively enhanced metabolic process in 'LP08' relative to 'Chiifu' grown in the light.

3.3 Identification of GSL pathway DEGs between 'Chiifu' and 'LP08'

We performed a search using the Basic Local Alignment Search Tool (BLAST) to identify GSL metabolic genes in the *B. rapa* genome, using sequence information for *Arabidopsis* GSL pathway genes collected from The *Arabidopsis* Information Resource (TAIR) website. In total, we collected 162 GSL pathway genes (19 TFs and 143 metabolic pathway genes) from the BRAD website (Supplementary Table S2). Among 162 *B. rapa* GSL pathway genes, 79 were DEGs between the 'Chiifu' and 'LP08' lines under light conditions (Supplementary Table S4). Among 79 GSL-related genes, 55 and 24 were up- and down-regulated, respectively, in 'LP08' compared to 'Chiifu' (Figure 2B and Supplementary Table S4). Interestingly, a majority of the up-regulated genes in 'LP08' (49 of 55 genes; 89%) were involved in the aliphatic GSL biosynthetic pathway (Figure 2C and Supplementary Table S4). This finding is in an agreement with the result showing significantly higher amounts of aliphatic GSLs in 'LP08' than 'Chiifu' (Figures 1A, B). Among the 24 down-regulated genes in 'LP08' compared to 'Chiifu', both

10.3389/fpls.2022.1067508

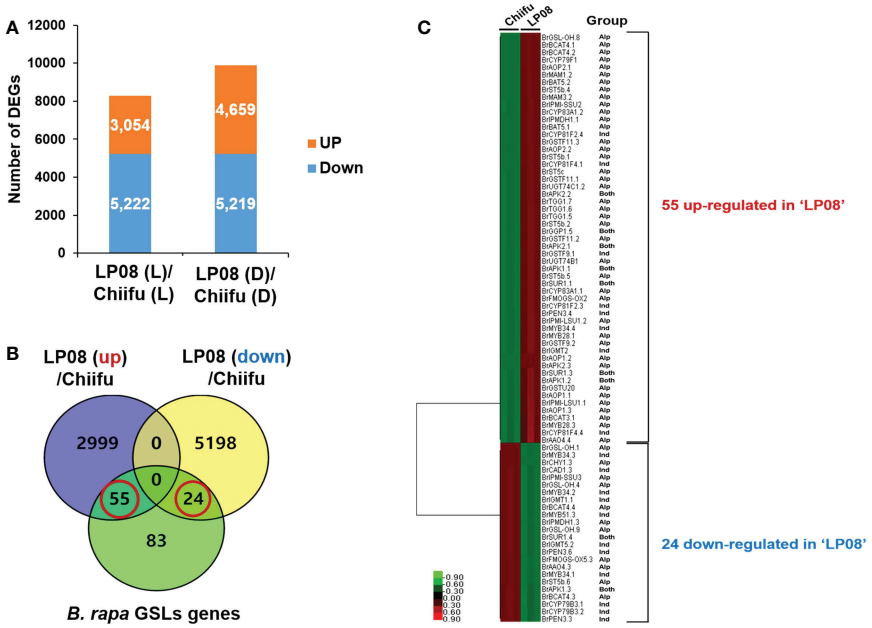


FIGURE 2
Identification of differentially expressed genes (DEGs) between 'Chiifu' and 'LP08' under light and dark conditions. **(A)** Bar graph showing the number of differentially expressed genes revealed by pairwise comparison between 'Chiifu' and 'LP08' grown under light (L) or dark (D) conditions. **(B)** Venn diagram showing 79 GSL pathway genes found among the 3,054 up-regulated and 5,222 down-regulated genes in 'LP08'. Out of these 79 genes, 55 and 24 were up- and down-regulated, respectively, in 'LP08' compared to 'Chiifu'. **(C)** Heatmap of 79 GSL pathway DEGs between 'Chiifu' and 'LP08'. Alp: aliphatic GSL pathway, Ind: indolic GSL pathway.

aliphatic and indolic GSL pathway genes were present in almost equal numbers (Figure 2C).

3.4 Expression of *B. rapa* MYB TFs regulating GSL biosynthesis

A subgroup of R2R3-type MYB TFs was reported to regulate GSL pathway genes. In total, 14 *B. rapa* MYB homologs (3 *BrMYB28* homologs, 1 *BrMYB29*, 3 *BrMYB34*, 3 *BrMYB51*, 2 *BrMYB118*, and 2 *BrMYB122*) were identified in BRAD (Supplementary Table S2). We examined the expression profiles of 14 *BrMYB* TFs between 'Chiifu' and 'LP08' grown under light and dark conditions using IGV. First, we analyzed six *BrMYBs* (*BrMYB28.1–3*, *BrMYB29*, and *BrMYB118.1–2*) involved in the aliphatic GSL pathway (Figure 3A). Expression levels of *BrMYB28.1* and *BrMYB28.3* were significantly higher in 'LP08' than 'Chiifu', whereas transcript levels of *BrMYB28.2* were similar between the two lines. Transcripts of *BrMYB29* and the two *BrMYB118* genes (*BrMYB118.1–2*) were not detected in young seedling plants grown under light or dark conditions. To validate the RNA-seq data, we conducted qRT-PCR analysis of aliphatic GSL pathway *BrMYB* genes (Supplementary Figure S4A). The results were similar to the RNA-seq data, confirming higher expression of *BrMYB28s* genes in 'LP08' than 'Chiifu'. Therefore, higher expression of *BrMYB28.1* and *BrMYB28.3* in 'LP08' likely contributes to the greater accumulation of aliphatic GSLs seen in 'LP08' than 'Chiifu'.

Among indolic pathway *BrMYB* TFs, the expression levels of eight *BrMYB* TFs (*BrMYB34.1–3*, *BrMYB51.1–3*, and *BrMYB122.1–2*) were compared between 'Chiifu' and 'LP08' under light and dark conditions (Figure 3B). Among these genes, the expression of *BrMYB34.1*, *BrMYB51.1*, and *BrMYB51.3* was strongly enhanced under dark conditions in both 'Chiifu' and 'LP08', whereas *BrMYB34.2* and *BrMYB34.3* were highly expressed under light conditions in the 'Chiifu' line but not the 'LP08' line (Figure 3B). While *BrMYB51.2* was expressed consistently regardless of genotype or light conditions, the two *BrMYB122* genes (*BrMYB122.1* and *BrMYB122.2*) and *BrMYB34.1* were expressed only in young seedlings under both light and dark conditions (Figure 3B). This suggested that individual *BrMYB* genes related to the indolic GSL pathway might regulate indolic GSL biosynthesis under certain environmental conditions or genotypes *via* diverse processes. Validation of RNA-seq data using qRT-PCR provided similar results (Supplementary Figure S4B). For example, *BrMYB34.1*, *BrMYB34.2*, and *BrMYB51.3* in the indolic GSL pathway were highly expressed in the dark, while *BrMYB34.3* was expressed in 'Chiifu' but not in 'LP08' (Figure 3B and Supplementary Figure S4B). Taken together, these results show that indolic GSL pathway *BrMYB* genes exhibit dynamic expression patterns in 'Chiifu' and 'LP08' under light and dark conditions, whereas aliphatic GSL pathway *BrMYB* genes are abundantly expressed in the light, particularly in the 'LP08' line.

Notably, the expression levels of aliphatic *BrMYB* genes were substantially higher than those of *BrMYB* genes in the indolic GSL pathway. The maximum intensity (y-axis of each IGV track view) of *BrMYB28.2* and *BrMYB28.3* was set to 5,000 and 2,500 read counts, respectively, whereas for *BrMYB51.2* and *BrMYB34.3* the values were 300 and 200, respectively. This difference indicates that aliphatic GSL pathway *BrMYBs* had relatively high expression compared to *BrMYB* genes in the indolic GSL pathway (Figure 3). This finding is in

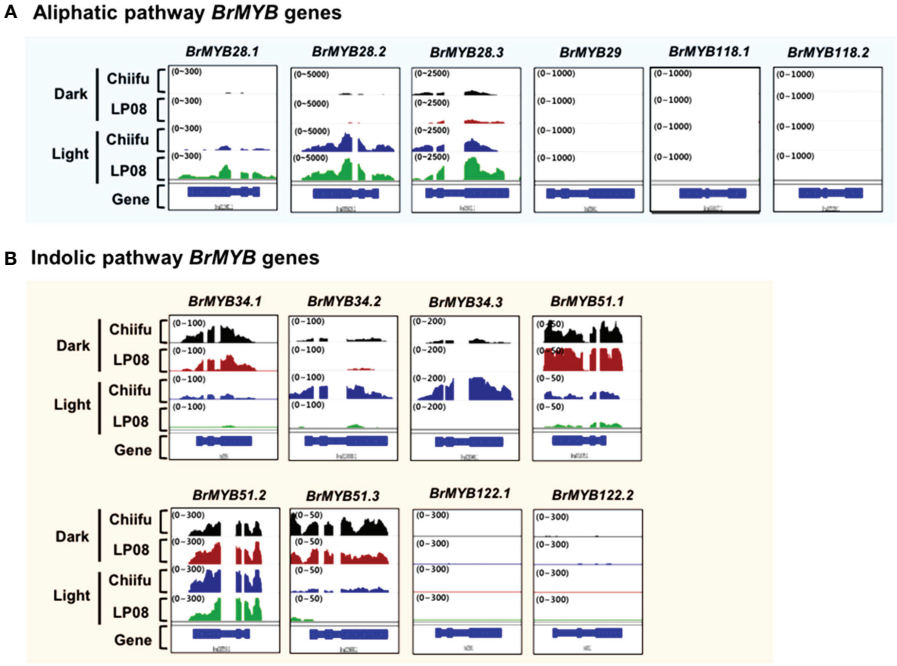


FIGURE 3
Expression profiles of *BrMYB* transcription factor **(TF)** genes between 'Chiifu' and 'LP08' grown under light and dark conditions. **(A)** Integrative genome browser (IGV) illustration of the expression profiles of six *BrMYB* genes involved in the aliphatic GSL pathway. Normalized read counts for each gene are indicated by black, brown, blue, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(B)** IGV illustration of the expression profiles of eight *BrMYB* genes involved in the indolic GSL pathway. Normalized read counts for each gene are indicated by black, brown, blue, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage, normalized using the total number of mapped reads, is indicated in parentheses at the top left corner of each track.

accordance with the level of aliphatic GSLs being substantially higher than the level of indolic GSLs in both the 'LP08' and 'Chiifu' lines.

3.5 Expression profiles of GSL metabolic genes

In general, genes involved in aliphatic and indolic GSL metabolism have been identified using *Arabidopsis* as a model plant (Supplementary Table S2). The GSL metabolic process can be grouped into five stages: 'side-chain elongation', 'core structure formation', 'secondary modification', 'co-substrate', and 'breakdown' (Supplementary Table S2). Among 143 *B. rapa* GSL metabolic genes (excluding TF genes), 72 (50%) were differentially expressed between the two lines grown in light conditions, with 55 being up-regulated and 24 down-regulated in 'LP08' compared to 'Chiifu' (Figure 2C and Supplementary Figure S4). A majority of up-regulated genes (46 of 55 in total; 84%) were involved in the aliphatic GSL pathway. Thus, it is plausible that higher expression of upstream *BrMYB* TF genes (*BrMYB28.1* and *BrMYB28.3*) in 'LP08' relative to 'Chiifu' results in up-regulation of downstream metabolic genes involved in aliphatic GSL biosynthesis (Figure 4).

For the 24 down-regulated genes in 'LP08' (up-regulated in 'Chiifu'), 11, 11 and 2 genes were related to the aliphatic, indolic, and both pathways, respectively (Figure 5 and Supplementary Table S4). This result indicates that indolic GSL metabolism is more diverse and complicated in these two lines than aliphatic GSL metabolism. We reasoned that the diverse composition of aliphatic and indolic GSLs in 'Chiifu' and 'LP08' might be derived in part from the dynamic expression of GSL metabolic genes. Taken together, our findings indicate that oilseed-type *B. rapa* 'LP08' had higher expression of *BrMYB28* TFs and aliphatic GSL metabolic genes compared to 'Chiifu', resulting in greater accumulation of aliphatic GSLs in 'LP08'. Among aliphatic GSL compounds, GNP was extraordinarily dominant. How the upregulation of *BrMYB28s* and downstream GSL metabolic genes drives specific accumulation of GNP among aliphatic GSL compounds remains unclear and requires further investigation.

Validation of the RNA-seq results was performed using qRT-PCR for 36 selected genes (24 aliphatic and 12 indolic pathway genes) involved in various stages of GSL metabolism. Among the 24 metabolic genes in the aliphatic GSL pathway, 23 were more abundantly expressed in 'LP08' than 'Chiifu'; the 1 exception was *BrGSL-OH.1* (Figure 5). Aliphatic GSL genes involved in the 'secondary modification' stage, such as *BrFMOGS-OX2*, *BrFMOGS-OX4*, *BrAOP2.1*, *BrAOP2.2*, *BrST5b.1*, *BrST5b.4*, and *BrST5c*, were more abundant in 'LP08' than 'Chiifu'. For the *GSL-OH.1* gene, the transcript level was much higher in 'Chiifu' than 'LP08', and was also higher in the dark than light condition. *GSL-OH.1* had low expression in 'LP08' under both light and dark conditions, in accordance with the RNA-seq results. In addition, between the light and dark conditions, most of the tested aliphatic GSL genes were more dominant in the light than dark, confirming the results of RNA-seq. In total, 12 genes related to the indolic GSL pathway were subjected to qRT-PCR and the results were similar to the RNA-seq data, confirming that

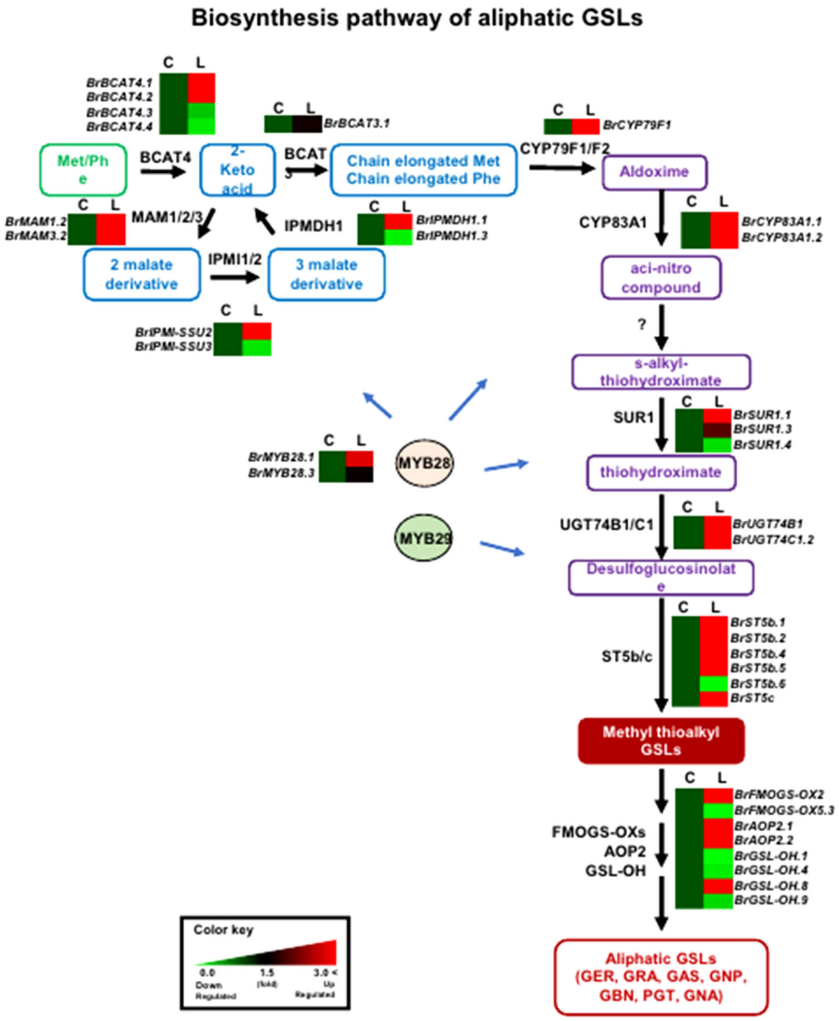


FIGURE 4
Comparison of expression profiles of GSL metabolic genes involved in aliphatic GSL biosynthesis. Schematic representation of aliphatic GSL metabolic pathways and expression patterns of aliphatic GSL pathway genes in 'Chiifu' and 'LP08'. The aliphatic GSL biosynthetic pathway begins with side-chain elongation of precursor amino acids such as methionine (Met) by catalytic enzymes such as BCAT4. Black letters between paths refer to genes associated with each step. *B. rapa* MYB TFs, such as *BrMYB28s* (*BrMYB28.1* and *BrMYB28.3*), boost the expression of aliphatic GSL metabolic genes. Compared to 'Chiifu', *BrMYB28s* and downstream GSL metabolic genes are more abundantly expressed in 'LP08'. For comparative heatmap analysis of 'Chiifu' and 'LP08', the level in 'Chiifu' was set to 1 and the relative level of each gene in 'LP08' is presented on the right side. C: Chiifu, L: LP08.

transcript patterns of indolic pathway genes differ between 'Chiifu' and 'LP08', under both light and dark conditions, more than aliphatic GSL genes (Figures 6, 7). Taken together, these results indicate that the oilseed-type *B. rapa* 'LP08' has higher expression of GSL metabolic genes, particularly aliphatic GSL pathway genes, than the vegetable-type 'Chiifu' line.

3.6 Identification of major QTLs associated with GSL content: Results of QTL mapping

Previously, we developed a mapping population comprised of F_5 151 RILs through the crossing of 'Chiifu' and 'LP08' (Kim et al., 2022). Among the 151 F_5 RILs, the seeds of 97 germinated successfully and grew well. These F_5 lines, and the parental lines, were used for extraction of GSLs (Supplementary Table S5). The amounts of individual GSL compounds from individual F_5 and parental lines were applied to linkage mapping based on 8,707 SNPs (Supplementary Figure S7). Although there are no peaks exceeding the LOD for total GSL, aliphatic GSL, and indole GSL, three QTL peaks exceeding the LOD (threshold = 5.906, 5.454, and 6.745, respectively) for GRA, GNP, and GBN were detected in chromosome A03 (Supplementary Figure S7; Figure 8A). these three peaks were corresponding to two peaks in upper arm regions (757 kb–787 kb) and the lower arm regions (approximately 1,950–2,210 kb) of total GSL and total Aliphatic GSL (Supplementary Figure S7). Within those regions, only two GSL pathway genes, *BrGSL-OH.1* (Bra022920) and *BrMYB28.1* (Bra012961), were located in the upper and lower arm QTL peak regions of chromosome A03, respectively. While *BrGSL-OH.1* was found at about 776 kb, *BrMYB28.1* was located at about 2,132 kb of chromosome 3. These two genes are both involved in the aliphatic GSL pathway, and no QTL candidate genes related to the indolic GSL pathway were detected in this study.

10.3389/fpls.2022.1067508

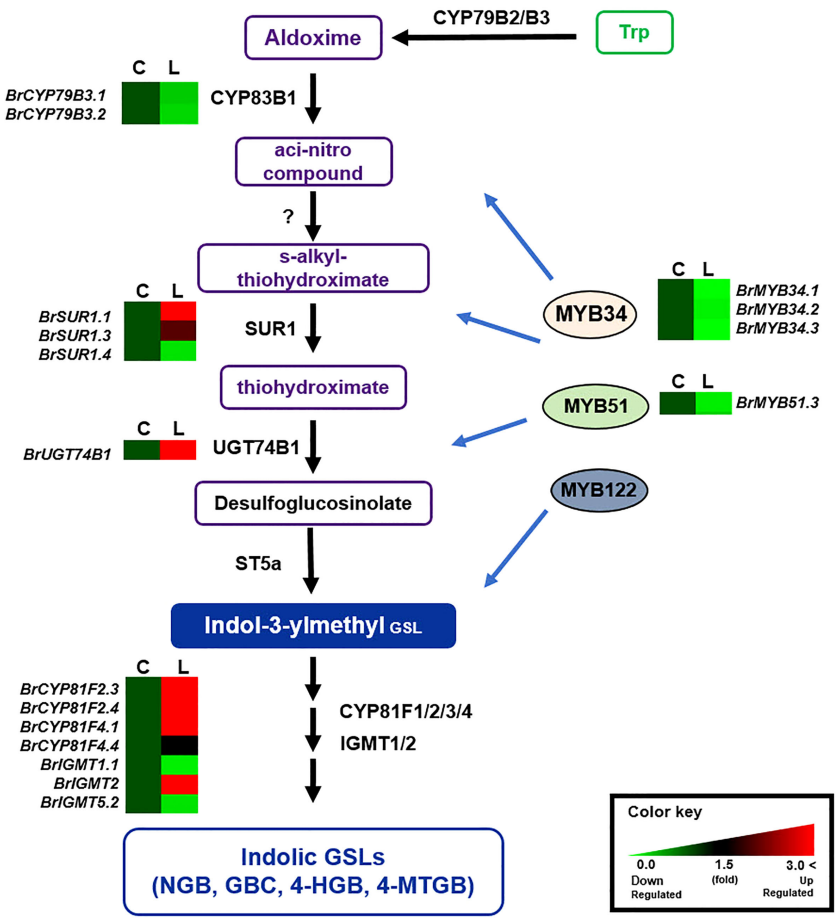


FIGURE 5
Comparison of expression profiles of GSL metabolic genes involved in indolic GSL biosynthesis. Schematic representation of indolic GSL metabolic pathways and expression patterns of indolic GSL pathway genes in 'Chiifu' and 'LP08'. The indolic GSL biosynthetic pathway begins with core structure formation from precursor amino acids, such as tryptophan (Trp), by catalytic enzymes such as CYP79B2. Black letters between paths refer to genes associated with each step. *B. rapa* MYB TFs such as BrMYB34s, BrMYB51s, and BrMYB122 boost the expression of indolic GSL metabolic genes. Expression profiles of *BrMYB34s, BrMYB51s, BrMYB122s*, and downstream indolic GSL metabolic genes, exhibited complex differences between 'Chiifu' and 'LP08'. For comparative heatmap analysis between 'Chiifu' and 'LP08', the level in 'Chiifu' was set to 1 and the relative level of each gene in 'LP08' is presented on the right side. C: Chiifu, L: LP08.

3.7 Comparison of genomic sequences of *BrMYB28.1* and *BrGSL-OH.1* between 'Chiifu' and 'LP08'

Two QTL genes (*BrMYB28.1* and *BrGSL-OH.1*) were included in the list of DEGs between the two lines (Figure 2C and Supplementary Table S4). *BrMYB28.1*, which was located in the highest QTL peak, was more abundant in 'LP08' than 'Chiifu', whereas *BrGSL-OH.1* showed the opposite pattern, with substantially higher expression in 'Chiifu' than 'LP08' (Figure 6). To elucidate the molecular details underlying the differential expression of these two genes, we cloned the genomic sequences of *BrMYB28.1* and *BrGSL-OH.1* from both the 'Chiifu' and 'LP08' lines. The results confirmed that both genes underwent significant genomic "context changes" from the promoter region to the 3′ downstream region between the two inbred lines (Figures 8B, C). For *BrGSL-OH.1*, several mutations including large deletions and insertions were identified in the promoter region of 'LP08' (Figure 8B). In the coding region, two point mutations were identified in the first exon (G to A) and second exon (C to T). In the 3′ downstream region, seven point mutations and two single-base insertions were identified.

Multiple mutations were also detected in the *BrMYB28.1* genomic region in 'LP08' compared to 'Chiifu'. Specifically, three point mutations, two deletions, and one 13-base insertion were identified in the promoter region of *BrMYB28.1* of 'LP08' (Figure 8C). In the coding region, two point mutations (A to T and A to C) were identified in the first intron and one point mutation (T to A) was found in the second exon (Figure 8C). In the 3′ downstream region, eight point mutations and three deletions were identified in 'LP08' compared to 'Chiifu'. As shown in Figure 3A and Supplementary Figure S4A, the expression of *BrMYB28.1* was elevated in 'LP08' compared to 'Chiifu', suggesting that genomic context changes in 'LP08' may enhance its transcriptional activity.

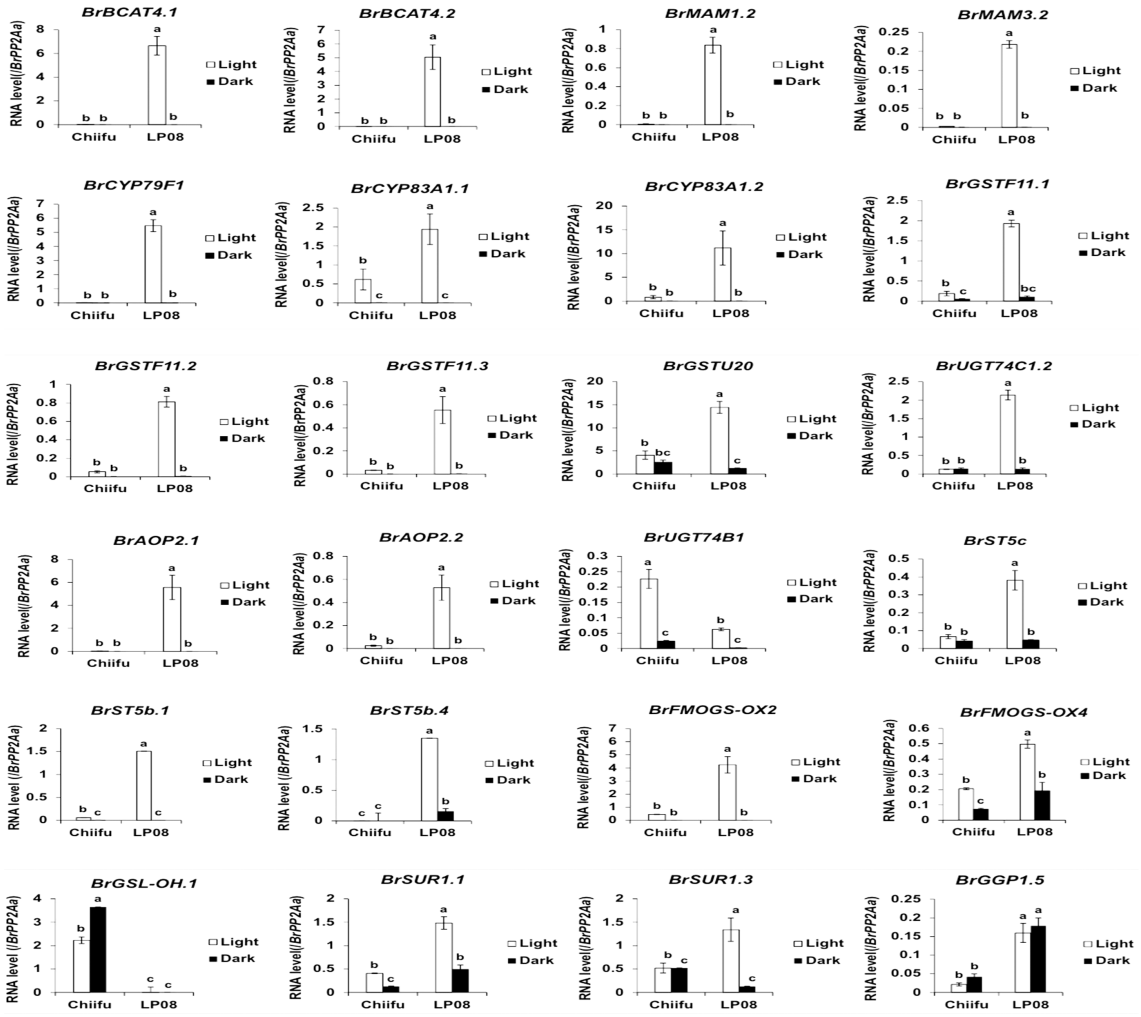


FIGURE 6
qRT-PCR analysis of 24 aliphatic GSL metabolic genes in 'Chiifu' and 'LP08' grown under light and dark conditions. Results of qRT-PCR analysis of 24 metabolic genes involved in aliphatic GSL biosynthesis. White and black bars indicate the expression level of each gene under light and dark conditions, respectively. Average values and standard deviations were calculated using the threshold cycle (Ct) values of three biological replicates. Significance was statistically determined using one-way analysis of variance (ANOVA) and Tukey's *post-hoc* test ($p < 0.05$), and indicated with different letters above bars.

4 Discussion

GSLs are secondary metabolites commonly synthesized in crop plants of family Brassicaceae, including important *Brassica* crops such as oilseed rape (*B. napus*), cabbage and broccoli (*B. oleracea*), and Chinese cabbage (*B. rapa*). GSLs play a major role as plant defensive compounds, and can affect the palatability and health value of edible crops (Grubb and Abel, 2006). Some GSLs and their degradation products have anti-carcinogenic and anti-oxidative activities in humans, and impart unique aromas and flavors to *Brassica* vegetables (Padilla et al., 2007). Due to their diverse roles in plant metabolism, animal nutrition, disease, and flavor, GSLs are a potential target for genetic manipulation and breeding for crop improvement. GSLs are derived from amino acids and can be divided into aliphatic (derived from Met, Leu, Ala, Ile, and Val), indolic (derived from Trp), and aromatic (derived from Phe and Tyr) classes according to their amino acid precursors (Grubb and Abel, 2006). Previous research using diverse *Brassica* subspecies reported that aliphatic GSL compounds accounted for the majority of GSLs, accounting for approximately 57–97% of the total GSL content (Kliebenstein et al., 2001; Bhandari et al., 2015; Seo et al., 2017). In accordance with a previous report (Seo et al., 2017), the total amounts of aliphatic GSLs were higher than those of indolic GSLs in this study (Supplementary Table S1). In addition, the amount of total GSLs (5,434.3 nmol/g) in 'LP08' was twice as high as in 'Chiifu' (2,869.9 nmol/g FW) (Figure 1B).

In addition to total GSL amounts, the composition of GSL compounds differed markedly between two inbred lines in terms of both aliphatic and indolic GSLs (Figures 1D, F). Previous studies have reported that GNP accounted for the majority of total GSLs in diverse *Brassica* species (Seo et al., 2017; Soundararajan et al., 2021). In this study, although 'LP08' (*B. rapa* ssp. *trilocularis*) is somewhat phenotypically different from vegetable-type *B. rapa* subspecies, GNP was the dominant aliphatic GSL compound in 'LP08' (Figure 1D). Meanwhile, 'Chiifu' (*B. rapa* ssp. *pekinensis*) a vegetable-type *B. rapa* line, had a diverse set of major GSLs, with three aliphatic GSLs (GRA, GNP, and GBN) exhibiting fairly similar proportions. This indicates that 'Chiifu' evolved to have a markedly

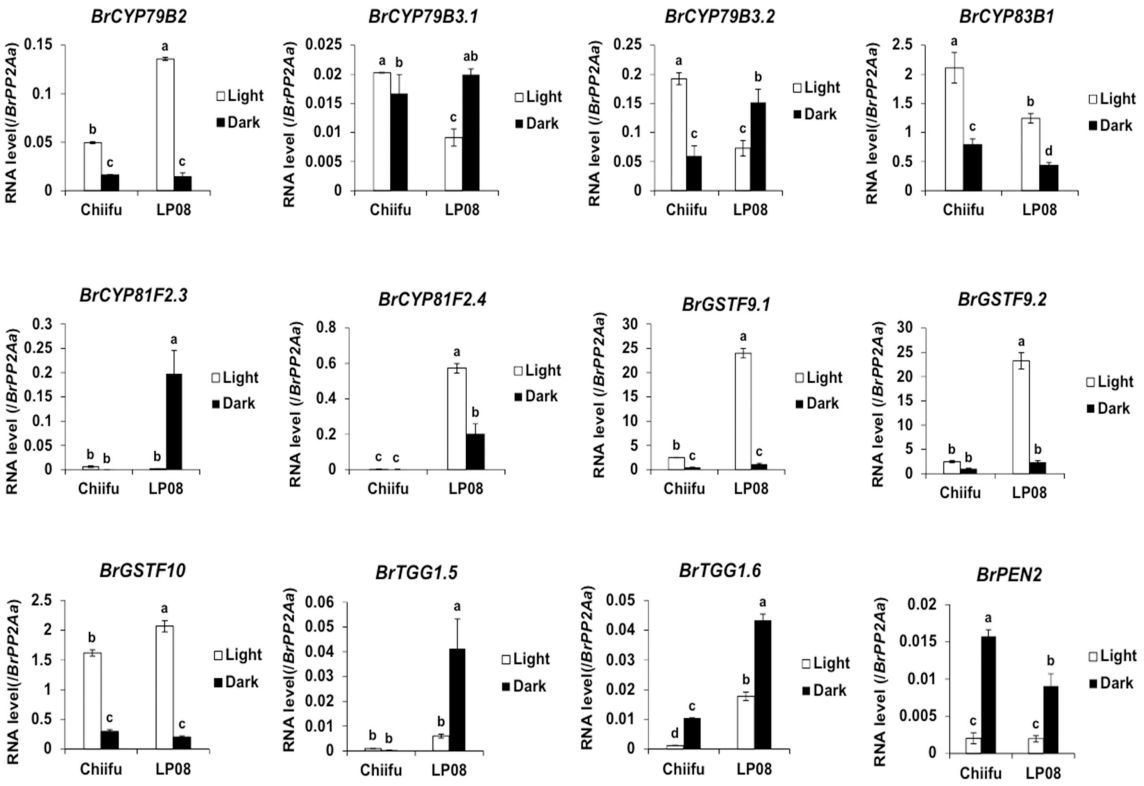


FIGURE 7
qRT-PCR analysis of 24 aliphatic GSL metabolic genes in 'Chiifu' and 'LP08' grown under light and dark conditions. Results of qRT-PCR analysis of 12 metabolic genes involved in indolic GSL biosynthesis. White and black bars indicate the expression level of each gene under light and dark conditions, respectively. Average values and standard deviations were calculated from the Ct values of three biological replicates. Significance was statistically determined using one-way ANOVA and Tukey's *post-hoc* test ($p < 0.05$), and indicated with different letters above bars.

different GSL profile from other *Brassica* subspecies. Further GSL profiling and transcriptomic analyses of diverse vegetable-type *B. rapa* plants, including 'Chiifu', might provide further insights into the divergent GSL profiles of *Brassica* subspecies.

In light condition, DEG analysis using RNA-seq between 'Chiifu' and 'LP08' found that 79 genes out of total 162 total GSL pathway genes were differentially expressed, showing 55 upregulated and 24 downregulated genes in 'LP08' in comparison to 'Chiifu', respectively (Figure 2B and Supplementary Table S4). Particularly, we noticed that a great portion of the upregulated genes in 'LP08' (49 of 55 genes; 89%) were related to the aliphatic GSL biosynthetic pathway (Figure 2C and Supplementary Table S4), explaining the reason why 'LP08' had significantly higher amounts of aliphatic GSLs than 'Chiifu' (Figures 1A, B). Additionally, expression levels of these 55 genes were also compared between light and dark condition, a majority of 55 genes were significantly reduced in dark samples of both 'LP08' and 'Chiifu' (Supplementary Figure S8A). This result suggest that aliphatic GSL biosynthesis is reduced in the absence of light, in an agreement with the previous study (Huseby et al., 2013). However, some of aliphatic GSL pathway genes including *BrAOP1.3*, *BrTGG1.6*, *BrST5b.5*, *BrAOP1.1*, and *BrAOP1.2* were rather upregulated in dark samples of both 'Chiifu' and 'LP08'. Functional role of these genes in dark condition needs further investigation. Furthermore, expression pattern of 24 downregulated genes in light condition were also analyzed in the RNA-seq dataset of dark condition. Compared to 55 upregulated genes, many of 24 downregulated genes were not significantly affected in the absence of light, less sensitively affected in the dark condition (Supplementary Figure S8B). Taken together, these data indicate that light-mediated signaling might play a positive effect on GSL biosynthesis pathway, particularly aliphatic GSL pathway in *B. rapa*.

Numerous environmental cues, including photoperiod (i.e. circadian rhythm) and other endogenous factors (i.e. phytohormones), affect GSL biosynthesis *via* modulation of GSL pathway genes (Bohinc and Trdan, 2012). Recently, we reported that a circadian clock component, *BrGI* (*B. rapa GIGANTEA*), is involved in the regulation of GSL biosynthesis in *B. rapa via* transcriptional modulation of GSL pathway genes (Kim et al., 2021). Crosstalk among diverse circadian components to coordinate daily GSL biosynthesis in *B. rapa* plants is an interesting topic for future research.

In QTL mapping, the second LOD peak was positioned around *BrGSL-OH.1* (Bra022920) (Figure 8A). Genomic sequence comparison betweeen 'Chiifu' and 'LP08' revealed that *BrGSL-OH.1* contained considerable mutations in 'LP08' compared to 'Chiifu'. For instance, large deletions and insertions in the promoter region, two point mutations like one in the first exon (G to A) and the other in the second exon (C to T), and seven point mutations and two single-base insertions in the 3′ downstream region (Figure 8B). Low expression of *BrGSL-OH.1* in 'LP08' might be attributed to severe mutational events, including large deletions and insertions in the promoter region of *BrGSL-OH.1*. We reasoned that these genomic mutations

10.3389/fpls.2022.1067508

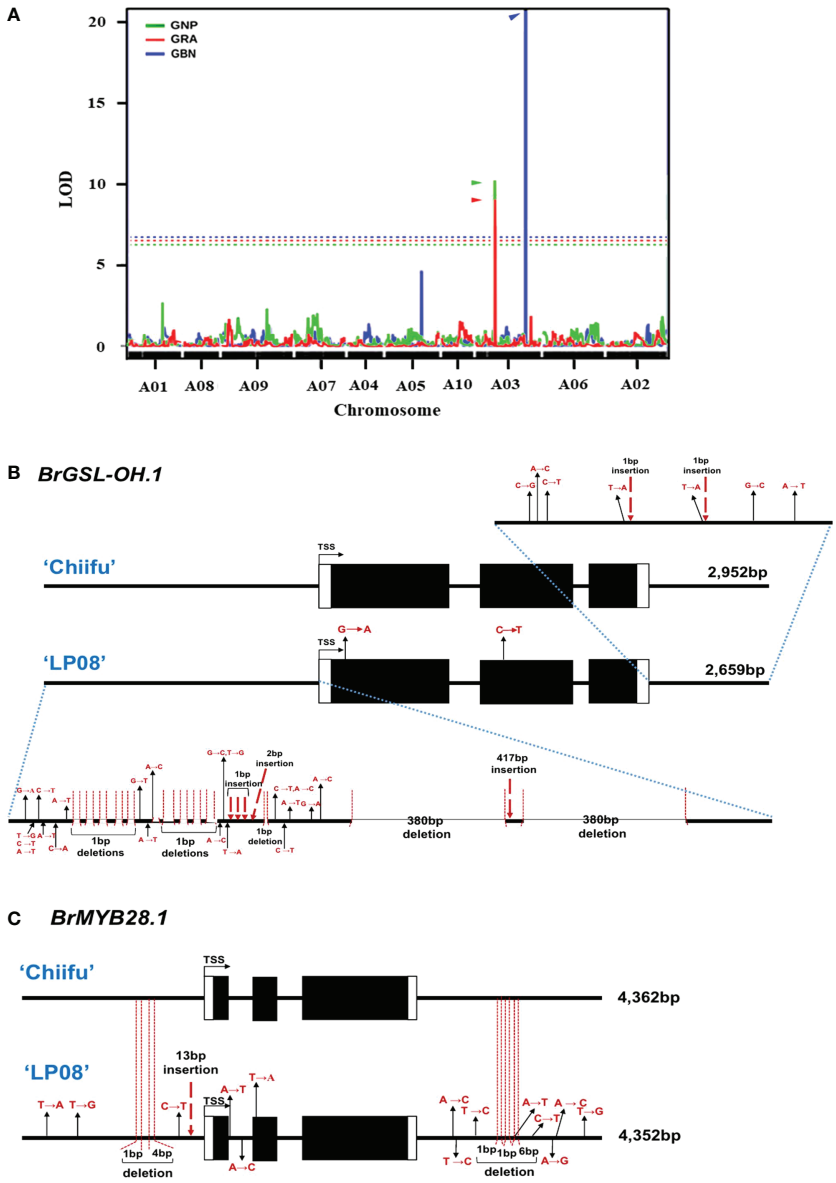


FIGURE 8
Identification of major QTL genes controlling GSL content in 'Chiifu' and 'LP08'. **(A)** Results of QTL mapping of 97 F_5 RILs and parental lines using HPLC data for GSL contents under long-day conditions. The X-axis represents the chromosomes (in order), and the Y-axis represents the limit of detection (LOD). Two peaks exceeding the LOD threshold for GRA and GNP were detected in the upper and lower arm regions for GBN in chromosome A03. The picture is made by overlapping three pictures to show the location of QTL peaks well. Green, red, and blue lines indicate the QTL peaks of GNP, GRA, and GBN, respectively and each threshold (dotted lines) is 5.454, 5.906, and 6.745. **(B)** Comparison of genomic structures of *BrGSL-OH.1* between 'Chiifu' and 'LP08'. The entire 2,952- and 2,659-bp genomic sequences of *BrGSL-OH.1* were amplified through PCR from 'Chiifu' (upper) and 'LP08' (lower), respectively. Comparison of the coding region of *BrGSL-OH.1* revealed two single-base substitutions (marked with black arrows) in the first and second exon of 'LP08'. Compared to *BrGSL-OH.1* in 'Chiifu', the promoter region of *BrGSL-OH.1* in 'LP08' had numerous single-base substitutions (marked with black arrows), along with 10 deletions (indicated with vertical dotted lines). Two large 380-bp deletions and one 417-bp insertion were identified in the promoter region of 'LP08'. In the 3′ downstream region, seven single-base substitutions (marked with black arrows) and two single-base deletions (indicated with vertical dotted lines) were identified in 'LP08' compared to 'Chiifu'. **(C)** Comparison of the genomic structures of *BrMYB28.1* between 'Chiifu' and 'LP08'. The entire 4,362- and 4,352-bp genomic sequences of *BrMYB28.1* were amplified by PCR from 'Chiifu' (upper) and 'LP08' (lower), respectively. Compared to 'Chiifu', four single-base substitutions (marked with black arrows), one single-base deletion and one 4-base deletion (indicated with vertical dotted lines), and one 13-bp insertion (marked with red dashed arrow) were identified in the promoter region of 'LP08'. In the coding sequence region, two single-base substitutions and one single-base substitution were detected in the first intron and second exon region, respectively, in 'LP08'. The 3′ downstream region of *BrMYB28.1* in 'LP08' contained eight single-base substitutions (marked with black arrows) and three deletions (indicated with vertical dotted lines).

might result in the loss of DNA elements required for active transcription of *BrGSL-OH.1*. *Arabidopsis GSL-OH* (Arabidopsis gene ID: AT2G25450) encodes a 2-oxoacid-dependent dioxygenase involved in the production of the aliphatic GSL compound PGT (Hansen et al., 2008). PGT has biological functions including toxicity against the nematode *Caenorhabditis elegans*, inhibition of seed germination, induction of goiter disease in mammals, and bitter taste in plants of the genus *Brassica*. In this study, the expression of *BrGSL-OH.1* was significantly higher in 'Chiifu' than 'LP08', such that 'Chiifu' is expected to have more PGT than 'LP08'. The amount of

PGT was higher in 'Chiifu' (30.8 nmol/g FW) than 'LP08' (21.4 nmol/g FW) (Supplementary Table S1). However, considering the dramatic up-regulation of *BrGSL-OH.1* in 'Chiifu', the quantitative difference between the two lines was subtle. One explanation for this is that the *B. rapa* genome may contain functionally redundant *BrGSL-OH* homologs. A BLAST search using the *Arabidopsis* GSL-OH sequence indicated that the *B. rapa* genome contains nine *B. rapa GSL-OH* homologs (named *BrGSL-OH.1–9*) (Supplementary Table S2). Among the nine homologs of *BrGSL-OH.1* to *BrGSL-OH.9*, four were differentially expressed between the two lines (Figure 2C and Supplementary Table S4). Three *BrGSL-OH* homologs (*BrGSL-OH.1*, *BrGSL-OH.4*, and *BrGSL-OH.9*) were expressed at higher levels in 'Chiifu' than 'LP08', whereas *BrGSL-OH.8* was more abundant in 'LP08' than 'Chiifu'. Therefore, the dynamic expression of multiple *BrGSL-OH* homologs in *B. rapa* may contribute to the moderate difference in PGT contents between the two inbred lines.

The highest LOD peak on the SNP-based linkage map was located in the lower arm region of chromosome A03 in the genomic region containing *BrMYB28.1* (Figure 8A). In *Arabidopsis*, a subgroup of MYB family TFs (*MYB28*, *MYB29*, and *MYB76*) have been reported to regulate the biosynthesis of aliphatic GSLs (Hirai et al., 2007; Gigolashvili et al., 2008; Sønderby et al., 2010a). In the *B. rapa* genome, three *MYB28* homologs [named *BrMYB28.1* (Bra012961), *BrMYB28.2* (Bra035929), *BrMYB28.3* (Bra029311)] and one *MYB29* homolog (*BrMYB29*, Bra005949) were identified (Supplementary Table S2). However, no *MYB76* homolog was found in the *B. rapa* genome. In this study, as well as *BrMYB28.1*, the expression of *BrMYB28.3* was also higher in 'LP08' than 'Chiifu', whereas *BrMYB28.2* had similar expression levels between the two lines. Another *BrMYB* genes, *BrMYB29* and *BrMYB118.1/2* were not expressed in our test irrespective of light condition (Figure 3A), suggesting that the *BrMYB29* and *BrMYB118.1/2* TFs do not contribute to aliphatic GSL biosynthesis, at least in seedling plants. *BrMYB29* and *BrMYB118s* may function in other developmental stages or under specific stress conditions. This hypothesis requires further investigation.

BrMYB28.1 was previously reported to have the highest transcript level among the three *BrMYB28* homologs in various organs of *B. rapa* (Seo et al., 2016). According to our RNA-seq results, however, *BrMYB28.1* expression was moderate compared to *BrMYB28.2* and *BrMYB28.3*. This discrepancy might result from differences in the growth conditions or developmental status of *B. rapa* plants. A

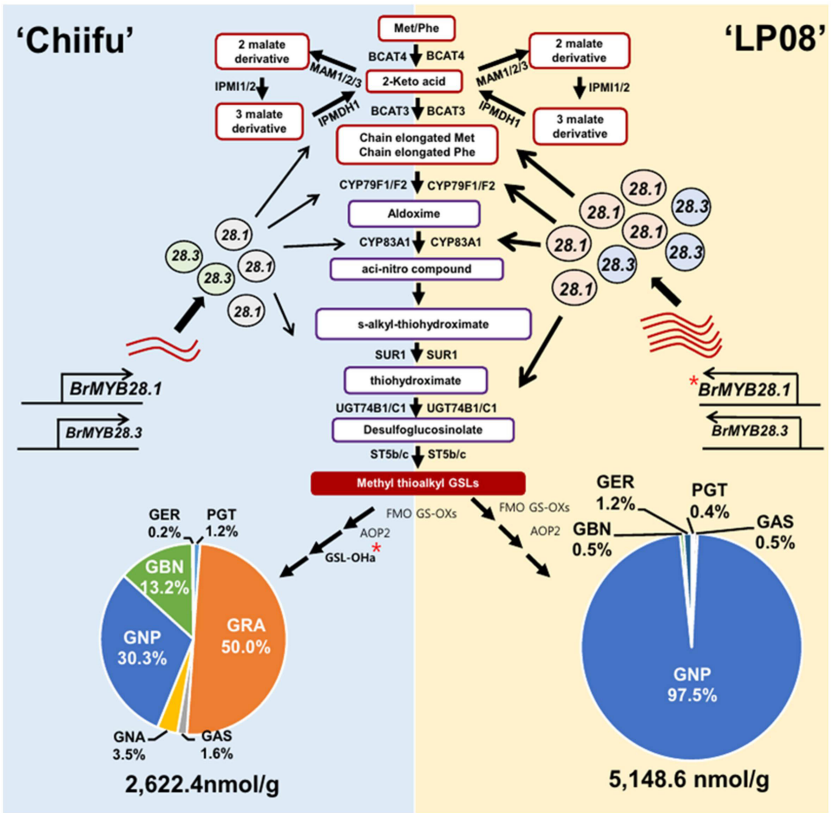


FIGURE 9
Schematic of the molecular mechanisms underlying the differing GSL profiles between vegetable-type 'Chiifu' and oilseed-type 'LP08'. The aliphatic GSL biosynthetic pathways of 'Chiifu' (left) and 'LP08' (right) are represented on blue and yellow backgrounds, respectively. Three BrMYB28 homologs play a positive role in the transcription of downstream GSL metabolic genes. Compared to 'Chiifu', 'LP08' had higher expression of *BrMYB28.1* (as well as *BrMYB28.3*), which stimulated downstream GSL metabolic genes. As a result, total amounts of aliphatic GSLs were significantly higher in 'LP08' than 'Chiifu'. In addition, *GSL-OH.1* was uniquely expressed in 'Chiifu' and not 'LP08', which may contribute to the differing compositional profiles of aliphatic GSLs between these two lines. 'Chiifu' has three major aliphatic GSL compounds accounting for large portions of its total aliphatic GSLs, whereas 'LP08' possessed a single GSL compound, GNP, accounting for 97.5% of the total aliphatic GSLs. Pie charts are presented as mean ± standard deviation **(SD)** (n = 3). Red asterisks (*) indicate two genes identified as major QTL candidates. PGT: progoitrin; GRE: glucoraphenin; GER: glucoerucin, GBN: glucobrassicanapin, GRA: glucoraphanin, GAS: glucoalyssin, GNP: gluconapin.

previous study reported that overexpression of *BrMYB28.1* resulted in elevated levels of aliphatic GSL compounds, indicating that BrMYB28.1 acts as a positive regulator of the GSL biosynthetic process in *B. rapa* plants, similar to the model plant *Arabidopsis* (Seo et al., 2016).

In this study, the expression of *BrMYB28.1* was higher in 'LP08' than 'Chiifu'. Regarding this observation, we reasoned that mutations in the promoter region of *BrMYB28.1* in 'LP08' may affect a DNA element required for the recruitment of a transcriptional repressor, leading to de-repression of the *BrMYB28.1* transcript level. Alternatively, mutations such as a 13-base insertion in the promoter region of *BrMYB28.1* may create a genomic context promoting the recruitment of transcriptional activators. These possibilities require further studies, including a promoter assay and complementation analysis to clarify the differences in profiles of GSL compounds between 'Chiifu' and 'LP08'. As a result of higher expression of *BrMYB28.1* in 'LP08', transcript levels of downstream GSL metabolic genes including *BrBCAT4.1/2*, *BrCYP79F1*, *BrCYP83.1/2*, *BrSUR1a*, *BrFMO GS-OX2*, and *BrAOP2.1/2* were also elevated in 'LP08' compared to 'Chiifu' (Figures 4, 6). Therefore, elevated expression of *BrMYB28.1* and its downstream genes likely greatly enhances the accumulation of GNP in 'LP08'. How elevated expression of *BrMYB28.1* and its downstream GSL metabolic genes leads to the overwhelming accumulation of GNP in 'LP08' remains unclear. Further analysis is needed to reveal the functional role of *BrMYB28.1* in the accumulation of GNP among aliphatic GSL compounds in *B. rapa* 'LP08' plants.

Based on the observations in this study, a conceptual model was proposed to describe the differing profiles of aliphatic GSLs between vegetable-type 'Chiifu' and oilseed-type 'LP08' (Figure 9). Precursor amino acids, such as Met and Phe, undergo a series of chemical modifications involving numerous enzymes related to GSL metabolism, including side chain elongation, core structure formation, and secondary modification. The *B. rapa* MYB genes *BrMYB28.1* and *BrMYB28.3* stimulate the expression of these GSL metabolic genes. Expression levels of *BrMYB28.1* and *BrMYB28.3* were higher in 'LP08' than 'Chiifu', resulting in higher expression of GSL metabolic genes related to aliphatic GSL biosynthesis in 'LP08'. In 'Chiifu', the QTL gene GSL-OH.1 was more abundant than in 'LP08', possibly causing differences in their profiles of aliphatic GSL compounds (abundant GNP, GRA, and GBN in 'Chiifu' vs. dominance by GNP in 'LP08').

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

JK and D-HK planned the experiments. JK, HK and HM prepared all plant materials. HM performed the molecular experiments. DC, SL, and AN performed the HPLC analysis. JK, HK, N-SK, and JJ developed the mapping population and performed the QTL analysis. HM and D-HK analyzed the RNA-seq data. HM, JK, and D-HK wrote the draft. D-HK revised and finalized the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Chung-Ang University Graduate Research Scholarship, awarded to HM in 2022, as well as the "New breeding technologies development Program of the Rural Development Administration, Republic of Korea" (project No. PJ01654101; funding awarded to JK) and NRF (2021R1F1A1047822; funding awarded to D-HK).

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.2022.1067508/full#supplementary-material

SUPPLEMENTARY FIGURE 1
UHPLC chromatogram of GSLs extracted from seedlings of 'LP08' (upper) and 'Chiifu' (bottom). 1, progoitrin (PGT); 2, glucoraphanin (GRA); 3, glucoalyssin (GAS); 4, gluconapoleiferin (GNA); 5, gluconapin (GNP); 6, 4-hydroxyglucobrassicin (4-HGB); 7, glucobrassicanapin (GBN); 8, glucoerucin (GER); 9, glucobrassicin (GBC); 10, 4-methoxyglucobrassicin (4-MTGB); 11, neoglucobrassicin (NGB); 12, gluconasturtiin (GNT).

SUPPLEMENTARY FIGURE 2
Identification of differentially expressed genes (DEGs) between 'Chiifu' and 'LP08'. **(A)** Multi-dimensional Scaling (MDS) plot showing that samples within a group had similar gene expression profiles and significant differences existed among groups. CL: Chiifu grown in the light; CD: Chiifu grown in the dark; LL: LP08 grown in the light; and LD: LP08 grown in the dark. Sample symbols are followed by the number of biological replicates (1–3). **(B)** Heatmap of the Pearson correlation matrix between samples used for pairwise comparisons. The color key was adjusted based on the log2-centered values to highlight differences. Dendrograms showing distances between repeated samples. **(C)** MA plot of DEGs between 'Chiifu' and 'LP08' grown under light (left) and dark (right) conditions. The X-axis represents average expression, logCPM (log_2-transformed counts per minute), and the Y-axis represents log_2-transformed fold-change values. Red dots indicate DEGs between 'Chiifu' and 'LP08'. Black dots indicate non-DEGs.

SUPPLEMENTARY FIGURE 3
Gene Ontology (GO) analysis of up- and down-regulated genes between 'Chiifu' and 'LP08' grown in the light condition. Top 10 functional categories of up-regulated (left) and down-regulated genes (right) in 'LP08' compared to 'Chiifu'. Categories related to the glucosinolates biosynthesis are shown in red, and red asterisks indicate significant categories. Y-axis indicate the percentage of detected category genes/total category genes which is automatically calculated by ShinyGO (v0.76) analysis tool.

SUPPLEMENTARY FIGURE 4
qRT-PCR analysis of two differentially expressed *BrMYB* genes between 'Chiifu' and 'LP08' grown under light and dark conditions. **(A)** Results of qRT-PCR analysis of two *BrMYB28s* (*BrMYB28.1* and *BrMYB28.3*) that exhibited differential expression between 'Chiifu' and 'LP08'. White and black bars indicate the expression levels of each gene under light and dark conditions, respectively. Average values and standard deviations were calculated from the Ct values of three biological replicates. **(B)** Results of qRT-PCR analysis of three *BrMYB34s* (*BrMYB34.1–34.3*) and a *BrMYB51.1* that exhibited differential expression between 'Chiifu' and 'LP08'. White and black bars indicate the expression levels of each gene under light and dark conditions, respectively. **(A, B)** Average values and standard deviations were calculated from the Ct values of three biological replicates. One-way ANOVA and Tukey's *post-hoc* test ($p < 0.05$) were used to determine statistically significant differences. Significance is represented by different letters above bars.

SUPPLEMENTARY FIGURE 5
Integrative genome browser (IGV) illustration of the expression profiles of total 42 aliphatic GSL biosynthetic genes differentially expressed between 'Chiifu' and 'LP08' grown under light and dark conditions. **(A)** IGV illustration of the expression profiles of 15 genes involved in the 'side-chain elongation' stage of aliphatic GSL biosynthesis. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(B)** IGV illustration of the expression profiles of 13 genes involved in the 'core structure formation' stage of aliphatic GSL biosynthesis. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(C)** IGV illustration of the expression profiles of 11 genes involved in the 'secondary modification' stage of aliphatic GSL biosynthesis. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(D)** IGV illustration of the expression profiles of three genes involved in the 'breakdown' stage of aliphatic GSL catabolism. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track.

SUPPLEMENTARY FIGURE 6
Integrative genome browser (IGV) illustration of the expression profiles of 19 total indolic GSL biosynthetic genes between 'Chiifu' and 'LP08' grown under light and dark conditions. **(A)** IGV illustration of the expression profiles of five genes involved in the 'core structure formation' stage of indolic GSL biosynthesis. Normalized read counts for each gene are indicated by black, brown, blue, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(B)** IGV illustration of the expression profiles of six genes commonly involved in the 'core structure formation' stages of both aliphatic and indolic GSL biosynthesis. Normalized read counts for each gene are indicated by black, brown, blue, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(C)** IGV illustration of the expression profiles of four genes involved in the 'secondary modification' stage of indolic GSL biosynthesis. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track. **(D)** IGV illustration of the expression profiles of four genes involved in the 'breakdown' stage of indolic GSL catabolism. Normalized read counts for each gene are indicated by black, blue, brown, and green colors for 'Chiifu' grown in the dark, 'LP08' grown in the dark, 'Chiifu' grown in the light, and 'LP08' grown in the light, respectively. Read coverage normalized to the total number of mapped reads is indicated in parentheses at the top left corner of each track.

SUPPLEMENTARY FIGURE 7
Results of QTL mapping of 97 F_5 RILs and parental lines using HPLC data for GSL contents under long-day conditions. **(A)** Results of QTL mapping for total GSLs, total indolic GSLs, and total aliphatic GSL. **(B)** Results of QTL mapping for individual aliphatic GSL compounds. Normalized read counts for each gene are indicated by black, brown, blue, and green colors for 'Chiifu' X-axis represents the chromosomes (in order), and the Y-axis represents the limit of detection (LOD). AlP and Ind means aliphatic GSL and indolic GSL. The red and blue asterisks indicate peaks at the same position, respectively.

SUPPLEMENTARY FIGURE 8
Expression profile of 77 DEGs (55 up- and 24 down-regulated genes under light and dark condition. **(A)** Expression heatmap of 55 'LP08'-upregulated GSL pathway genes under light and dark condition. A majority of 55 'LP08' up-regulated genes were significantly downregulated in both 'Chiifu' and 'LP08' dark samples. Downregulated genes in dark condition were indicated with right side blue vertical bar. Upregulated genes in the dark condition were indicated with right side red vertical bar. **(B)** Expression heatmap of 24 'LP08'-downregulated GSL pathway genes under light and dark condition. Many of 24 'LP08'-downregulated genes were not significantly affected in the absence of light, less sensitively affected in the dark condition when compared to the expression profile of 55 'LP08' up-regulated genes.

References

Agerbirk, N., and Olsen, C. E. (2012). Glucosinolate structures in evolution. *Phytochemistry* 77, 16–45. doi: 10.1016/j.phytochem.2012.02.005

Bellostas, N., Kachlicki, P., Sørensen, J. C., and Sørensen, H. (2007). Glucosinolate profiling of seeds and sprouts of b. oleracea varieties used for food. *Sci. Hortic. (Amsterdam).* 114, 234–242. doi: 10.1016/j.scienta.2007.06.015

Bhandari, S. R., Jo, J. S., and Lee, J. G. (2015). Comparison of glucosinolate profiles in different tissues of nine brassica crops. *Mol.* 20, 15827–15841. doi: 10.3390/MOLECULES200915827

Bohinc, T., and Trdan, S. (2012). Environmental factors affecting the glucosinolate content in brassicaceae. *J. Food Agric. Environ.* 10, 357–360.

Dixon, R. A. (2001). Natural products and plant disease resistance. *Nature* 411, 843–847. doi: 10.1038/35081178

Gigolashvili, T., Berger, B., Mock, H. P., Müller, C., Weisshaar, B., and flügge, U.I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in arabidopsis thaliana. *Plant J.* 50, 886–901. doi: 10.1111/j.1365-313X.2007.03099.x

Gigolashvili, T., Engqvist, M., Yatusevich, R., Müller, C., and flügge, U.I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in arabidopsis thaliana. *New Phytol.* 177, 627–642. doi: 10.1111/j.1469-8137.2007.02295.x

Grubb, C. D., and Abel, S. (2006). Glucosinolate metabolism and its control. *Trends Plant Sci.* 11, 89–100. doi: 10.1016/j.tplants.2005.12.006

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Han, N., Ku, K. M., and Kim, J. (2019). Postharvest variation of major glucosinolate and their hydrolytic products in brassicoraphanus 'BB1.'. *Postharvest Biol. Technol.* 154, 70–78. doi: 10.1016/J.POSTHARVBIO.2019.04.011

Hansen, B. G., Kerwin, R. E., Ober, J. A., Lambrix, V. M., Mitchell-Olds, T., Gershenzon, J., et al. (2008). A novel 2-Oxoacid-Dependent dioxygenase involved in the formation of the goiterogenic 2-Hydroxybut-3-enyl glucosinolate and generalist insect resistance in arabidopsis. *Plant Physiol.* 148, 2096–2108. doi: 10.1104/PP.108.129981

Herten, K., Hestand, M. S., Vermeesch, J. R., and Van Houdt, J. K. J. (2015). GBSX: A toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinf.* 16, 1–6. doi: 10.1186/S12859-015-0514-3/

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of arabidopsis myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104

Hopkins, R. J., Van Dam, N. M., and Van Loon, J. J. A. (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu. Rev. Entomol.* 54, 57–83. doi: 10.1146/annurev.ento.54.110807.090623

Huseby, S., Koprivova, A., Lee, B. R., Saha, S., Mithen, R., Wold, A. B., et al. (2013). Diurnal and light regulation of sulphur assimilation and glucosinolate biosynthesis in arabidopsis. *J. Exp. Bot.* 64, 1039–1048. doi: 10.1093/JXB/ERS378

Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52, 39. doi: 10.1186/s40659-019-0246-3

Jia, C. G., Xu, C. J., Wei, J., Yuan, J., Yuan, G. F., Wang, B. L., et al. (2009). Effect of modified atmosphere packaging on visual quality and glucosinolates of broccoli florets. *Food Chem.* 114, 28–37. doi: 10.1016/J.FOODCHEM.2008.09.009

Kakizaki, T., Kitashiba, H., Zou, Z., Li, F., Fukino, N., Ohara, T., et al. (2017). A 2-Oxoglutarate-Dependent dioxygenase mediates the biosynthesis of glucoraphasatin in radish. *Plant Physiol.* 173, 1583–1593. doi: 10.1104/PP.16.01814

Keck, A. S., and Finley, J. W. (2004). Cruciferous vegetables: Cancer protective mechanisms of glucosinolate hydrolysis products and selenium. *Integr. Cancer Ther.* 3, 5–12. doi: 10.1177/1534735403261831

Kim, J. A., Kim, J. S., Hong, J. K., Lee, Y.-H., Lee, S. I., Jeong, M. J., et al. (2017). Development of a marker system to discern the flowering type in brassica rapa crops. *J. Plant Biotechnol.* 44, 438–447. doi: 10.5010/JPB.2017.44.4.438

Kim, N. S., Kim, S. J., Jo, J. S., Lee, J. G., Lee, S. I., Kim, D. H., et al. (2021). The brgi circadian clock gene is involved in the regulation of glucosinolates in chinese cabbage. *Genes (Basel)* 12, 1664. doi: 10.3390/GENES12111664/S1

Kim, S., Kim, J. A., Kang, H., and Kim, D. H. (2022). A premature stop codon in BrFLC2 transcript results in early flowering in oilseed-type brassica rapa plants. *Plant Mol. Biol.* 108, 241–255. doi: 10.1007/S11103-021-01231-Y

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/GB-2013-14-4-R36

Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., et al. (2001). Genetic control of natural variation in arabidopsis glucosinolate accumulation. *Plant Physiol.* 126, 811–825. doi: 10.1104/pp.126.2.811

Koprivova, A., and Kopriva, S. (2014). Molecular mechanisms of regulation of sulfate assimilation: First steps on a long road. *Front. Plant Sci.* 5. doi: 10.3389/FPLS.2014.00589/

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 94 (9), 357–359. doi: 10.1038/nmeth.1923

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

Li, Y., Sawada, Y., Hirai, A., Sato, M., Kuwahara, A., Yan, X., et al. (2013). Novel insights into the function of arabidopsis R2R3-MYB transcription factors regulating aliphatic glucosinolate biosynthesis. *Plant Cell Physiol.* 54, 1335–1344. doi: 10.1093/PCP/PCT085

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

Padilla, G., Cartea, M. E., Velasco, P., de Haro, A., and Ordás, A. (2007). Variation of glucosinolates in vegetable crops of brassica rapa. *Phytochemistry* 68, 536–545. doi: 10.1016/J.PHYTOCHEM.2006.11.017

Sønderby, I. E., Burow, M., Rowe, H. C., Kliebenstein, D. J., and Halkier, B. A. (2010a). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in arabidopsis. *Plant Physiol.* 153, 348–363. doi: 10.1104/PP.109.149286

Sønderby, I. E., Geu-Flores, F., and Halkier, B. A. (2010b). Biosynthesis of glucosinolates - gene discovery and beyond. *Trends Plant Sci.* 15, 283–290. doi: 10.1016/j.tplants.2010.02.005

Sønderby, I. E., Hansen, B. G., Bjarnholt, N., Ticconi, C., Halkier, B. A., and Kliebenstein, D. J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PloS One* 2, e1322. doi: 10.1371/JOURNAL.PONE.0001322

Saldanha, A. J. (2004). Java Treeview–extensible visualization of microarray data. *Bioinformatics* 20, 3246–3248. doi: 10.1093/BIOINFORMATICS/BTH349

Schonhof, I., Krumbein, A., and Brückner, B. (2004). Genotypic effects on glucosinolates and sensory properties of broccoli and cauliflower. *Nahrung - Food* 48, 25–33. doi: 10.1002/food.200300329

Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., and Binder, S. (2006). BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in arabidopsis. *Plant Cell* 18, 2664–2679. doi: 10.1105/TPC.105.039339

Seo, M. S., Jin, M., Chun, J. H., Kim, S. J., Park, B. S., Shon, S. H., et al. (2016). Functional analysis of three BrMYB28 transcription factors controlling the biosynthesis of glucosinolates in brassica rapa. *Plant Mol. Biol.* 90, 503–516. doi: 10.1007/S11103-016-0437-Z/

Seo, M. S., Jin, M., Sohn, S. H., and Kim, J. S. (2017). Expression profiles of BrMYB transcription factors related to glucosinolate biosynthesis and stress response in eight subspecies of brassica rapa. *FEBS Open Bio* 7, 1646–1659. doi: 10.1002/2211-5463.12231

Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., et al. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in arabidopsis. *Plant J.* 47, 10–24. doi: 10.1111/J.1365-313X.2006.02767.X

Soundararajan, P., Park, S. G., Won, S. Y., Moon, M. S., Park, H. W., Ku, K. M., et al. (2021). Influence of genotype on high glucosinolate synthesis lines of brassica rapa. *Int. J. Mol. Sci.* 22, 7301. doi: 10.3390/IJMS22147301/S1

Sun, B., Liu, N., Zhao, Y., Yan, H., and Wang, Q. (2011). Variation of glucosinolates in three edible parts of Chinese kale (Brassica alboglabra bailey) varieties. *Food Chem.* 124, 941–947. doi: 10.1016/J.FOODCHEM.2010.07.031

Thinh Nguyen, V. P., Stewart, J., Lopez, M., Ioannou, I., and Allais, F. (2020). Glucosinolates: Natural occurrence, biosynthesis, accessibility, isolation, structures, and biological activities. *Mol.* 25, 4537. doi: 10.3390/MOLECULES25194537

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/BIB/BBS017

Vallejo, F., Tomás-Barberán, F. A., Gonzalez Benavente-García, A., and García-Viguera, C. (2003). Total and individual glucosinolate contents in inflorescences of eight broccoli cultivars grown under various climatic and fertilisation conditions. *J. Sci. Food Agric.* 83, 307–313. doi: 10.1002/JSFA.1320

Vallejo, F., Toms-Barbern, F. A., and Garca-Viguera, C. (2002). Potential bioactive compounds in health promotion from broccoli cultivars grown in Spain. *J. Sci. Food Agric.* 82, 1293–1297. doi: 10.1002/JSFA.1183

Verhoeven, D. T. H., Goldbohm, R. A., Van Poppel, G., Verhagen, H., and Van Den Brandt, P. A. (1996). Epidemiological studies on brassica vegetables and cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 5, 733–748.

Verpoorte, R., and Memelink, J. (2002). Engineering secondary metabolite production in plants. *Curr. Opin. Biotechnol.* 13, 181–187. doi: 10.1016/S0958-1669(02)00308-7

Wang, J., Gu, H., Yu, H., Zhao, Z., Sheng, X., and Zhang, X. (2012). Genotypic variation of glucosinolates in broccoli (Brassica oleracea var. italica) florets from China. *Food Chem.* 133, 735–741. doi: 10.1016/j.foodchem.2012.01.085

Wittstock, U., and Halkier, B. A. (2002). Glucosinolate research in the arabidopsis era. *Trends Plant Sci.* 7, 263–270. doi: 10.1016/S1360-1385(02)02273-2

Xu, C. J., Guo, D. P., Yuan, J., Yuan, G. F., and Wang, Q. M. (2006). Changes in glucoraphanin content and quinone reductase activity in broccoli (Brassica oleracea var. italica) florets during cooling and controlled atmosphere storage. *Postharvest Biol. Technol.* 42, 176–184. doi: 10.1016/j.postharvbio.2006.06.009

Zhang, J., Liu, Z., Liang, J., Wu, J., Cheng, F., and Wang, X. (2015). Three genes encoding AOP2, a protein involved in aliphatic glucosinolate biosynthesis, are differentially expressed in brassica rapa. *J. Exp. Bot.* 66, 6205–6218. doi: 10.1093/JXB/ERV331

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fpls.2023.1111418

OPEN ACCESS

EDITED BY
Xiaodong Yang,
Yangzhou University, China

REVIEWED BY
Pilar Soengas,
Biological Mission of Galicia (CSIC), Spain
Xiaonan Li,
Shenyang Agricultural University, China

*CORRESPONDENCE
Yihua Liu
18225126866@163.com
Jingjing Chen
chjjmao@163.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 29 November 2022
ACCEPTED 13 February 2023
PUBLISHED 22 February 2023

CITATION
Chen B, Liu Y, Xiang C, Zhang D, Liu Z, Liu Y and Chen J (2023) Identification and *in vitro* enzymatic activity analysis of the *AOP2* gene family associated with glucosinolate biosynthesis in Tumorous stem mustard (*Brassica juncea* var. *tumida*). *Front. Plant Sci.* 14:1111418. doi: 10.3389/fpls.2023.1111418

Identification and *in vitro* enzymatic activity analysis of the *AOP2* gene family associated with glucosinolate biosynthesis in Tumorous stem mustard (*Brassica juncea* var. *tumida*)

Bing Chen, Yu Liu, Chunfang Xiang, Dandan Zhang, Zhuoyu Liu, Yihua Liu* and Jingjing Chen*

School of Life Advanced Agriculture Bioengineering, Yangtze Normal University, Chongqing, China

The major enzyme encoded by the glucosinolate biosynthetic gene *AOP2* is involved in catalyzing the conversion of glucoiberin (GIB) into sinigrin (SIN) in Brassicaceae crops. The *AOP2* proteins have previously been identified in several Brassicaceae species, but not in Tumorous stem mustard. As per this research, the five identified members of the *AOP2* family from the whole genome of *Brassica juncea* named *BjuAOP2.1-BjuAOP2.5* were found to be evenly distributed on five chromosomes. The subcellular localization results implied that BjuAOP2 proteins were mainly concentrated in the cytoplasm. Phylogenetic analysis of the AOP2 proteins from the sequenced *Brassicaceae* species in BRAD showed that *BjuAOP2* genes were more closely linked to *Brassica carinata* and *Brassica rapa* than *Arabidopsis*. In comparison with other *Brassicaceae* plants, the *BjuAOP2* members were conserved in terms of gene structures, protein sequences, and motifs. The light response and hormone response elements were included in the *BjuAOP2* genes' cis-regulatory elements. The expression pattern of *BjuAOP2* genes was influenced by the different stages of development and the type of tissue being examined. The BjuAOP2 proteins were used to perform the heterologous expression experiment. The results showed that all the five BjuAOP2 proteins can catalyze the conversion of GIB to SIN with different catalytic activity. These results provide the basis for further investigation of the functional study of *BjuAOP2* in Tumorous stem mustard glucosinolate biosynthesis.

KEYWORDS

***Brassica juncea*, glucosinolate, *BjuAOP2*, expression pattern, prokaryotic expression, activity analysis**

1 Introduction

Brassica juncea var. *tumida* (tumorous stem mustard, TSM) is an important vegetable crop of the Brassica genus of the Cruciferae family that originated in China. The enlarged fleshy stems of TSM are the raw material for Fuling Preserved Szechuan Pickle and can be

eaten fresh. Approximately 130 glucosinolates (GSLs), an important class of secondary metabolites that are sulfur- and nitrogen-rich and found primarily in cruciferous plants, have been identified (Blažević et al., 2019; Nguyen et al., 2020). GSLs undergo hydrolysis reactions under their degradation enzyme, myrosinase, to produce products such as isothiocyanates, thiocyanates, and acetonitrile (Chen et al., 2020). The afore-mentioned products are involved in anti-cancer and anti-bacterial (Augustine and Bisht, 2015; Soundararajan and Kim, 2018) processes, resistance to herbivore feeding, pathogenic microbial infestation (Clay et al., 2009; Li et al., 2014; Chen et al., 2020), and the formation of specific flavors in cruciferous vegetables (Engel et al., 2006). The biosynthetic precursors of GSLs include amino acids such as methionine, tryptophan, phenylalanine, and leucine. These amino acids form side chains of different lengths through an extension pathway. GSLs can be categorized as aromatic, aliphatic, and indole types as per the derivation sources of the side chains (Halkier and Gershenzon, 2006). The highest content in TSM is 2-propenyl glucosinolate (sinigrin, SIN), the major aliphatic GSLs present in *B. juncea* (Li et al., 2011). TSM with high SIN content had a pronounced spicy taste (Mazumder et al., 2016). Therefore, revealing the synthesis process and molecular regulation mechanism of SIN is crucial to improve the flavor of TSM and optimize the GSL fraction in TSM.

Aliphatic GSL biosynthesis in *Arabidopsis thaliana* comprises three independent steps: elongation of the precursor amino acid-based side chain, core structure formation, and side chain secondary modification (Grubb and Abel, 2006; Sønderby et al., 2010). The current research indicates the involvement of almost 15 transcription factors and 64 structural genes in GSL biosynthesis (Li et al., 2018; Harun et al., 2020; Mitreiter and Gigolashvili, 2021). GSL-Elong is the key site in the side chain elongation step. Three genes, MAM1 (methylthioalkylmalate synthases 1), MAM2, and MAM3, located in a tandem arrangement at this locus (Kroymann et al., 2001; Kroymann et al., 2003; Textor et al., 2007). The enzyme known as MAM1 and MAM2 catalyzes the condensation reaction of the first three elongation cycles whereas MAM3 mediate the condensation reaction of all six elongation cycles (Textor et al., 2007). The functional polymorphism of the MAM gene results in differences in the structure of the GSL among different ecotypes of *Arabidopsis thaliana* (Wittstock and Halkier, 2002). The cytochrome P450 homolog *CYP79* gene family, *CYP83* gene family, and *UDP-glucosyltransferase 74B1* (UGT74B1) are key genes in the core structure formation step (Grubb et al., 2004; Mikkelsen et al., 2004). The flavin-containing monooxygenases (GSL-OX), 2-oxoglutarate-dependent dioxygenases (GSL-AOP), and 2-oxoacid-dependent dioxygenase (GSL-OH) sites during modification of the GSL side chain determine the diversity of GSL species (Kliebenstein et al., 2001a; Hansen et al., 2008). GSL-AOP includes two closely linked genes, GSL-ALK and GSL-OHP (Kliebenstein et al., 2001a). GSL-ALK (*AOP2*) and GSL-OHP (*AOP3*) genes, respectively, catalyze the transformation of methylsulfinyl GSLs into alkenyl GSLs and hydroxyalkyl GSLs (Kliebenstein et al., 2001b). In addition, the enzymes encoded by the *AOP2* gene can catalyze the conversion of beneficial GSL glucoraphanin (GRA) with anticancer activity to gluconapin (GNA) and glucoiberin (GIB) to SIN, thus affecting the taste and flavor of cruciferous vegetables.

The *AOP* gene family in *Arabidopsis thaliana* includes *AOP1*, *AOP2*, and *AOP3* (Kliebenstein et al., 2001a). The *AOP1* gene may be the ancestor of the other two genes, but its function needs to be further characterized (Neal et al., 2010). *AOP* genes greatly vary in their expression patterns in various *Arabidopsis thaliana* ecotypes and under different culture conditions. The *AOP2* is expressed in L*er* (Landsberg), but *AOP3* is absent, while Cvi (Cape Verde Islands), depicts the expression of *AOP3*, but *AOP2* is absent (Kliebenstein et al., 2001a). In addition, an ecotype in which both genes are expressed simultaneously has not been found yet. Studies in *Arabidopsis thaliana* and *Brassica napus* suggested that inactivation of the *AOP2* gene promotes the accumulation of large amounts of GRA in plants to produce higher amounts of sulforaphane with anticancer activity (Neal et al., 2010). Conversely, increased expression of the *AOP2* gene causes the GNA and SIN content to increase in plants (Neal et al., 2010; Liu et al., 2012). *In vitro* activity analysis of *Arabidopsis thaliana AOP2* protein by Kliebenstein et al. suggested that the GRA could be converted to GNA through an induced *AOP2* protein solution (Kliebenstein et al., 2001b).

Brassica, a genus of economic and nutritional importance in the Brassicaceae family, consists of six species involving three diploids: *B. rapa* (2n=20, AA), *B. oleracea* (2n=18, CC), and *B. nigra* (2n=16, BB). Mutual hybridization and continuous selective evolution among the three yielded three allotetraploids: *B. carinata* (2n=34, BBCC), *B. juncea* (2n=36, AABB), and *B. napus* (2n=38, AACC). The number and expression pattern of *AOP* genes in Brassica vary from species to species. For example, the *AOP2* gene in both *B. rapa* and *B. oleracea* has three homologs (Gao et al., 2004; Liu et al., 2014). Two and four homologs of the *AOP2* gene were found in the genomes of cabbage mustard and oil-mustard, respectively (Augustine and Bisht, 2015; Wu et al., 2017). The *BoAOP2* gene in *B. oleracea* can be expressed normally to catalyze the degradation of GRA into GNA (Zheng et al., 2022). However, there is a nonfunctional *AOP2* allele in broccoli with a two-base deletion on the exon, which in turn influences the accumulation of glucoraphanin products (Li and Quiros, 2003). Liu et al. reported that *B. oleracea* contained another non-functional *BoAOP2*. The translation termination caused by the premature termination codon functioning of this gene also caused the accumulation of glucoraphanin products (Liu et al., 2014). The presence of three alleles of *AOP2* in *B. rapa*, all with catalytic activity, was found to be tissue expression-specific (Wang et al., 2011). Augustine et al. documented a significant reduction in the amount of SIN in transgenic plants than in wild-type plants by constructing an *AOP2* silencing vector in oil mustard (Augustine and Bisht, 2015). A recently conducted study in pennycress revealed that the SIN content was significantly reduced in the wild-type F_2 population versus the *AOP2* mutant (Chopra et al., 2020). Therefore, the *AOP2* gene is a bridge between beneficial and deleterious GSLs. Although several relevant studies have been conducted on this gene in *B. oleracea* and *B. rapa*, the mechanism that underlies its functions in TSM has not been studied in detail. Therefore, elucidating the mechanism of *AOP2* genes in GSL synthesis and degradation remains a great challenge due to the complexity of the Brassica crop genome (Malhotra and Bisht, 2020).

In this study, five genes homologous to *Arabidopsis thaliana AOP2* were cloned using the allotetraploid crop TSM (2n=36, AABB) as plant material. In addition, the protein's physicochemical properties, gene structure, phylogenetic tree, promoter cis-acting elements, subcellular localization, and gene expression were comprehensively analyzed. The gene functions were initially verified using *in vitro* experiments. This study revealed the role of individual *BjuAOP2* genes in the process of SIN synthesis in TSM and the expression differences among different copies. Overall, the data obtained in this research acts as a foundation for revealing the molecular mechanism of SIN synthesis regulation by *BjuAOP2* genes in *Brassica juncea* and offer novel insights into improving the GLS fraction of TSM at the molecular level.

2 Materials and methods

2.1 Sample sources

The TSM high-generation selfing line B186 was sown in the Yangtze Normal University squash trial site in the fall of 2021 after germination. Per the normal growth requirements of TSM, field management was performed during growth. The leaves of B186 were collected at the 4-leaf stage for *BjuAOP2* cloning, and the expanded tumorous stems of B186 were collected at 4, 7, 10, 13, 16, 19, and 22 weeks after sowing to analyze the expression pattern of *BjuAOP2* at different developmental stages. In addition, roots, stems, leaves, tumorous stems, flowers, and siliques of B186 were collected at the shooting stagegib for analysis of the gene's tissue expression pattern. Each sample had three biological replicates that were collected and placed in liquid nitrogen and kept at -80°C for subsequent use.

2.2 Identification of the *BjuAOP2* gene family

The protein sequence of the *Arabidopsis thaliana AOP2* gene was retrieved online from TAIR (http://www.arabidopsis.org/). Subsequently, a BlastP search was performed in the Brassica genome database (http://brassicadb.cn/#/BLAST/) to search for amino acid sequences and nucleic acid sequences of TSM *AOP2* gene family members. In addition, HMMsearch (Johnson et al., 2010) was used to identify the 2OG-FeII_Oxy (PF03171) and DIOX-N(PF14226) structural domains for all possible *AOP2* genes. Finally, conserved structural domain confirmation was performed through NCBI-CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) in NCBI to exclude sequences without typical structural domains.

2.3 Cloning and sequence analysis of the *BjuAOP2* gene

Total RNA of TSM leaves was isolated with Tiangen RNA Extraction Kit TRNzol Universal Reagent (Tiangen Biotech, Beijing, China) and with a subsequent reversal to single-stranded cDNA using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech, Beijing, China). The designing of the primers (Supplementary Table S1) and amplification of the full-length CDS sequences using Primer premier 6.0 software was the next step, using the *BjuAOP2* gene sequence obtained from the stem mustard database as a template. The PCR reaction system and reaction procedure were performed according to the instructions of TransStart FastPfu DNA Polymerase from TransGen, and the PCR product recovery and purification were performed according to the instructions of the TransGen Gum Recovery Kit (EG101). In addition, the PCR-recovered products were homologously cloned with the vector pTF101-GFP using the homologous recombination method of the pEASY-Basic Seamless Cloning and Assembly Kit (TransGen Biotech, Beijing, China) from TransGen. Subsequently introducing the recombinant vector into *E. coli* receptor DH5α, plasmids were extracted from 12 positive clones for validation and sequencing. The protein sequences encoding the *BjuAOP2* gene obtained by sequencing were subjected to multiple-sequence alignment with the *AOP2* protein sequences of *Arabidopsis thaliana* using DNAStar 7.1 software. The CDS length and amino acid numbers of the *BjuAOP2* gene were obtained from the clone sequencing results. Data such as chromosome position was obtained from the reference genome. Isoelectric points and molecular weights were obtained from the pI/Mw calculation tool on the website ExPASy (http://www.expasy.org/tools/). The prediction of subcellular localization was obtained using online tool BUSCA (http://busca.biocomp.unibo.it/).

2.4 Phylogenetic tree construction of the *AOP* gene family in Brassicaceae crops

Genomic data of 19 sequenced cruciferous species were obtained from The Brassicaceae Genome Resource (http://www.tbgr.org.cn) (Liu et al., 2022). The protein sequence of the *Arabidopsis thaliana AOP* gene was accessed at the TAIR website as a reference sequence and the *AOP* family gene sequence was obtained using the Quick Find Best Homology tool in TBtools software (Chen et al., 2020). Clustal W was employed to carry out the multiple-sequence alignment of *AOP* family proteins from sequenced species in the Brassicaceae family in MAGE 11 software. Subsequently, utilizing the neighbor-joining (NJ) method phylogenetic trees were drawn, with the Bootstrap method set to 1000 (Tamura et al., 2021). The online tool, GSDS (http://gsds.cbi.pku.edu.cn/index.php) was employed for mapping the intron-exon structure pattern of the *AOP* gene. The analysis of the conserved motif of the gene was performed using the MEME (http://meme-suite.org/) online tool. The upstream promoter sequences (1.5-kb) of each gene of the identified *AOP* family were accessed at the Brassica genome website (http://brassicadb.org) with the promotor-bound cis-acting elements were examined by the software PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html). Schematic diagrams of all the above analysis results were illustrated utilizing the software Domain Illustrator software (http://dog.biocuckoo.org/) (Ren et al., 2009).

2.5 Protein subcellular localization of *BjuAOP2*

Protein subcellular localization of *BjuAOP2* was performed by transient expression in tobacco leaf epidermal cells. *Agrobacterium tumefaciens* GV3101 containing the *ProCAMV35S::GFP::BjuAOP2* vector plasmid was activated and enlarge-cultivated. Subsequently, the organisms were collected, resuspended in a resuspension solution (OD600 almost 0.5), left for 2-3 h, and introduced into the lower epidermis of 3-4 weeks old tobacco leaves with a syringe. Observations were made after 3 days of transformation. The leaves transformed with empty pTF101-GFP were used as control. In addition, laser confocal microscopy (Nikon, Tokyo, Japan) was utilized for the observation of GFP fluorescence with excitation light at 488 nm and emission light at 510 nm. Furthermore, chloroplast autofluorescence showed excitation light at 640 nm and emission light at 675 nm.

2.6 Prokaryotic expression and *in vitro* enzyme activity assay of the *BjuAOP2* gene

The design of the primers and amplification of full-length CDS sequences were executed through Primer premier 6.0 software, using the *BjuAOP2* gene sequence obtained from the stem mustard database as a template (Supplementary Table S2). The pET-32a expression vector (Novagen. Madison, WI, USA) was constructed using the full-length CDS sequence of the gene through homologous recombination with the pEASY-Basic Seamless Cloning and Assembly Kit (TransGen Biotech, Beijing, China). Subsequently, plasmids were extracted from 12 positive clones for validation and sequencing. The correctly sequenced recombinant plasmids for the prokaryotic expression of the gene were transformed into *Escherichia coli* strain Transetta (DE3). The recombinant pET vector contains the T7 promoter and capable of expressing a fusion protein containing thioredoxin. The *E. coli* DE3 strain containing prokaryotic expression of the recombinant plasmid was incubated in LB medium at 37°C while also being shaken until OD600 was approximately = 0.6. Then, a final concentration of 0.5 mM was obtained by adding IPTG, and the expression of recombinant protein was induced by shaking the bacteria overnight at 16°C. The bacteria were isolated after induction by centrifuging the *E. coli* broth for five minutes at 6000 rpm/min. These bacteria were then suspended in PBS and placed on an ultrasonic cell crusher for 3 min at 100MHz (5-second on-10-second off cycle). The obtained cell-disrupted solution was centrifuged (12000 rpm/min and 20 min) for supernatant collection. Subsequently, the supernatant was purified by fusion protein according to the instruction procedures of Ni IDA Beads 6FF Kit (Changzhou Smart-Lifesciences Biotechnology, Changzhou, China). Finally, the purified fusion protease was detected by SDS-PAGE electrophoresis.

Kliebenstein et al. (2001a) validated the *Arabidopsis thaliana AOP2* protein by *in vitro* experiments. In this procedure, the same method was used to validate the function of *BjuAOP2* protein in *Arabidopsis thaliana*, with a slight modification of the enzyme activity assay. The procedure was as follows: 400 μL of purified *BjuAOP2* protein solution, 10 mM ascorbate, 15 mM α-ketoglutarate, 200 mM sucrose, 200 μM FeSO4, and 100 μL GIB were added to five 10 mL centrifuge tubes (total reaction volume: 4 mL). A control 10 mL centrifuge tube was added with 400 μL ddH2O instead of purified *BjuAOP2* protein solution, and the other components remained unchanged. Subsequently, the six centrifuge tubes were placed on a shaker at 28°C (110 rpm) for 4 hours. The reaction solution was purified by desulfurization and subjected to high-performance liquid chromatography (HPLC) to analyze the GSL fraction

2.7 GSL extraction and HPLC analysis

GSLs were extracted and assayed per the prior procedure (He et al., 2002). First, 1 mL of the enzyme-activated reaction solution was transferred to a DEAE Sephadex A-25 column for overnight desulfurization at room temperature using lipase sulfate (Sigma, E.C. 3.1.6.) and eluted twice with 750 μL of deionized water. The filtration of the eluent was carried out through a 0.22 μm Hydrophilic polyethersulfone (PES) Syringe Filter (Shanghai Anpel) for HPLC analysis. The elute, the external standard 3-GIB, and SIN were analyzed by HPLC.

2.8 Quantitative real-time fluorescence analysis

Total RNA was obtained from different tissues and developmental stages of tuberous mustard through TRNzol Universal Total RNA Extraction Reagent from Tiangen Biotech (Beijing, China). cDNA was synthesized based on the CDS sequences obtained by homologous recombinant cloning and sequencing using Primer premier 6.0. The afore-mentioned software designed real-time fluorescence quantification primers (Supplementary Table S3). The reaction system was as follows: 2×SuperReal PreMix Plus 10 μL; upstream and downstream primers, 0.5 μL each; cDNA template, 2 μL; ddH2O, 7 μL. In addition, a three-step amplification method was used: 95°C pre-denaturation for 5 min; 95°C for 30 s, 58°C for 30 s, and 72°C for 30 s (40 cycles). Dissociation curve: 65-95°C (0.5°C increase per cycle) for 5 s (1 cycle). The entire fluorescent quantitative PCR reaction was performed on a Roche Light Cycler 480 instrument. The proportionate expression of the five *BjAOP2* was calculated using the $2^{-\Delta\Delta CT}$ calculation method with actin as an internal reference gene (Livak and Schmittgen, 2001). Technical and biological replicates, three each were utilized in each experiment.

2.9 Statistical analysis

One-way analysis of variance (ANOVA), significance analysis, and Duncan's multiple-comparison were performed on the

experimental data using SPSS analysis software, with a significance threshold of p-value < 0.05.

3 Results

3.1 Identification and cloning of *BjuAOP2* gene family members

BLAST alignment of the *Arabidopsis thaliana AOP2* gene protein sequences with the TSM database was performed. Subsequently, the 2OG-FeII_Oxy (PF14226) structural domain was identified for all possible *AOP2* genes. In total, five members of the gene family distributed in different chromosomes, named *BjuAOP2.1-BjuAOP2.5*, were identified from the TSM genome database (Table 1). The CDS sequences of the genes mentioned above were obtained by PCR using the cDNA obtained by reverse transcription as a template. *BjuAOP2* genes have CDS sequences in the length range of 1284-1428 bp, with encoding amino acid numbers between 427-475 aa, molecular weight ranging from 46.8 to 55.2 kDa, and isoelectric point variation ranging from 4.77 to 5.66. The optimal prediction of subcellular localization is nucleus.

Multiple-sequence alignment with the *Arabidopsis thaliana AOP2* gene (Figure 1) depicted that the *BjuAOP2* gene family's five members contain the key structural domains, DIOX-N and 2 OG-Fell-Oxy, and the sequences in this range are highly conserved. However, the protein sequence similarity of the middle part of these five members is low, leading to the divergence of gene functions among the members.

3.2 Phylogenetic, gene structure, motif and cis-acting regulatory elements analysis of the *BjuAOP2* gene family

An *AOP* gene evolutionary tree was constructed for TSM and 18 other cruciferous crops that have been sequenced to further understand the evolutionary origins of the five members of the *BjuAOP2* gene family in TSM (Figure 2; Supplementary Table S4). The *AOP* gene family is divided into two major categories: (1) *AOP1s* consisting of 33 *AOP1* amino acid sequences; (2) *AOP2s* and *AOP3s* consisting of 25 *AOP2* amino acids and 3 *AOP3* amino acid sequences. In addition, small branches (N ≤ 3) formed by the genes are all derived from different Brassicaceae species, and the small branches further form large branches in the form of clusters containing two and three genes. Members of the TSM *AOP* gene family are more similar to those in *B. carinata* and *B. rapa* than those in *Arabidopsis thaliana*. The TSM *BjuAOP2* gene has the closest affinity to *B. rapa*, and *B. nigra*, which is consistent with the evolutionary origin doctrine in U's Triangle Brassica Species (Cheng et al., 2016).

The gene structure diagram from the *AOP* gene family (Figure 3A) revealed that the *AOP* family is conserved in the number of exons, 53 of the 61 members containing three exons. Despite the similar exon length of most genes, the intron lengths differed significantly. The *AOP* family was analyzed for conserved motifs using the MEME online tool (Figure 3B). In total, 15 conserved motifs of 8-41 aa in length were contained in the *AOP* gene (Figure 3C; Supplementary Table S5). In addition, the structural domains of 15 motifs were analyzed using NCBI-CDD online software. The motif 10,4,5 and 7 together form the DIOX_N domain. And the motif 2, 6, 8 and 1 together form the 2OG-FeII_Oxy domain structural of the *AOP* gene family.To further elucidate the functional element that influences expression of *AOP* genes, the promoter sequences were analyzed using the PlantCARE database to identify cis-regulatory elements in the promoter region. Nineteen types of stress- and hormone-related cis-acting regulatory elements were detected in the promoters of *AOP* genes (Figure 4; Supplementary Table S6). All 61 *AOP* genes contained element related to light. The MYB bingding site were detected in 47. Hormone-related elements (MeJA-responsiveness, salicylic acid responsiveness, zein metabolism regulation, abscisic acid responsiveness, auxin-responsive element, gibberellin-responsive element) were detected in more than 45 members of *AOP* genes. Other results are shown in the Figure 4.

3.3 Subcellular localization of the protein encoded by the *BjuAOP2* gene

The green fluorescent protein (GFP) fusion transient expression vector containing the *BjuAOP2* gene was injected into tobacco leaves and observed by laser confocal microscopy (Figure 5). Except for

TABLE 1 Basic information of the *BjuAOP2* genes.

Gene ID	Rename	chromosome	Genome location		Gene length/ bp	Number of amino acids/ aa	Molecular weight/ kDa	isoelectric point	Subcellular localization
			Start	End					
BjuVB05G45660	BjuAOP2.1	B05	46966364	46970953	1299	432	47.6	4.77	nucleus
BjuVA09G02070	BjuAOP2.2	A09	1311307	1313569	1320	439	47.9	4.89	nucleus
BjuVA02G29850	BjuAOP2.3	A02	17116725	17119697	1284	427	46.8	4.95	nucleus
BjuVB08G37450	BjuAOP2.4	B08	28215201	28217478	1428	475	52.2	5.66	nucleus
BjuVA03G30000	BjuAOP2.5	A03	14204864	14211046	1323	440	48.3	4.90	nucleus

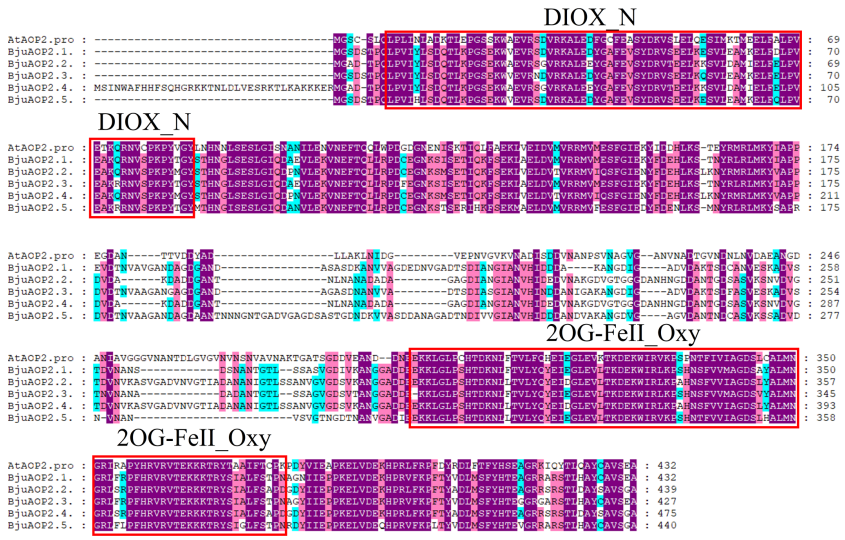


FIGURE 1
Multiple-sequence alignment of Tumorous stem mustard BjuAOP2 and AtAOP2 protein sequence. Multiple alignments were performed using the MEGA11 software. The red solid boxes represent the structural domains of 2OG-FeII_Oxy and DIOX-N. The purple, blue and pink shading indicates 100%, 80% and 60% conserved percent, respectively.

BjuAOP2.4, where a green fluorescent signal was observed only in the nucleus, the other four family members were observed to fluoresce in both the cytoplasm and nucleus. Thus, *BjuAOP2.1*, *BjuAOP2.2*, *BjuAOP2.3*, and *BjuAOP2.5* were expressed in the cytoplasm and nucleus, and *BjuAOP2.4* was expressed only in the nucleus, consistent with the prediction results of subcellular localization.

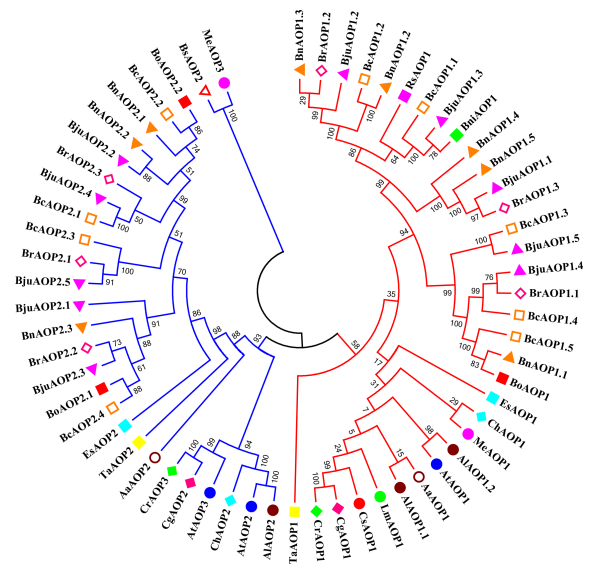


FIGURE 2
Phylogenetic tree of AOP genes from Brassicaceae species. The phylogenetic tree was built using the neighbor-joining (NJ) method. The unrooted tree was generated using Clustal W in MEGA11 *via* AOP amino acid sequences from 19 sequenced Brassicaceae species, including: *Arabis alpina* (*Aa.*), *Arabidopsis lyrate* (*Al.*), *Arabidopsis thaliana* (*At.*), *Brassica carinata* (*Bc.*), *Brassica juncea* (*Bju.*), *Brassica napus* (*Bn.*), *Brassica nigra* (*Bni.*), *Brassica oleracea* (*Bo.*), *Brassica rapa* (*Br.*), *Boechera stricta* (*Bs.*), *Capsella grandiflora* (*Cg.*), *Cardamine hirsute* (*Ch.*), *Capsella rubella* (*Cr.*), *Eutrema salsugineum* (*Es.*), *Lepidium meyenii* (*Lm.*), *Microthlaspi erraticum* (*Me.*), *Raphanus sativus* (*Rs.*), and *Thlaspi arvense* (*Ta.*). Genes from each species were labeled with different colors. The numbers at the nodes represent bootstrap percentage values.

3.4 Expression pattern of the *BjuAOP2* gene

The expression patterns of five *BjAOP2* genes in various tissue parts of B186 and various developmental stages of the tumorous stem were analyzed using real-time PCR (Figure 6). Expression of *BjuAOP2* genes was detected in roots, stems, tumorous stems, leaves, flowers and siliques. *BjuAOP2* genes had significantly different expression patterns in different tissues (Figure 6A). *BjuAOP2.2* were more highly expressed in flowers than in other tissues examined (Figure 6A), whereas *BjuAOP2.1*, *BjuAOP2.3* and *BjuAOP2.4* were more highly expressed in stems than in other tissues (Figure 6A). *BjuAOP2.5* were more highly expressed in tumorous stems than in other tissues (Figure 6A). Analysis of the expression pattern of the *BjuAOP2* gene at different stages of tumorous stem development revealed different expression patterns with the expansion of TSM (Figure 6B). The expression of *BjuAOP2.1*, *BjuAOP2.2* and *BjuAOP2.4* peaked at 7 weeks after sowing before falling back to basal levels. *BjuAOP2.3* and *BjuAOP2.5* showed low expression level throughout the whole growth period.

3.5 *In vitro* enzyme activity analysis of *BjuAOP2*

BjuAOP2 prokaryotic expression vector was constructed using PET-32a. The protein expression was induced by IPTG, and the *BjuAOP2* fusion protein was obtained by the His-tag purification

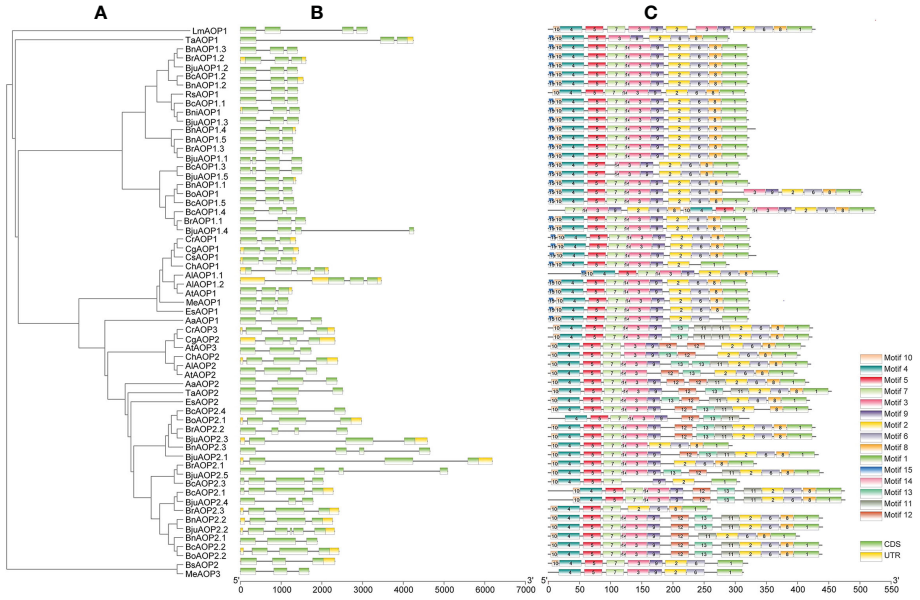


FIGURE 3
Gene structures and motifs of 61 *AOP* genes identified in *B juncea* and other sequenced Brassicaceae crops. **(A)** unrooted phylogenetic relationships among the AOP protein sequences. **(B)** exon-intron organization of corresponding AOP gene. Exons are represented by green boxes, while introns are represented by gray lines. **(C)** conserved motifs of AOP genes. Colored boxes indicated conserved motifs and gray lines represent non-conserved sequences. The length of motifs in each protein was showed proportionally.

column. *In vitro* enzymatic activity analysis of five BjuAOP2 fusion proteins was performed using GIB as substrate. As shown in Figure 7, all five BjuAOP2 proteins successfully catalysed the conversion of GIB to SIN. The standards GIB and SIN were detected at 3.3 min and 4.6 min, where only GIB was detected in the control group. However, both GIB and SIN were detected in the protein elutes of all five members. Therefore, all five *BjuAOP2* proteins of TSM have catalytic activity.

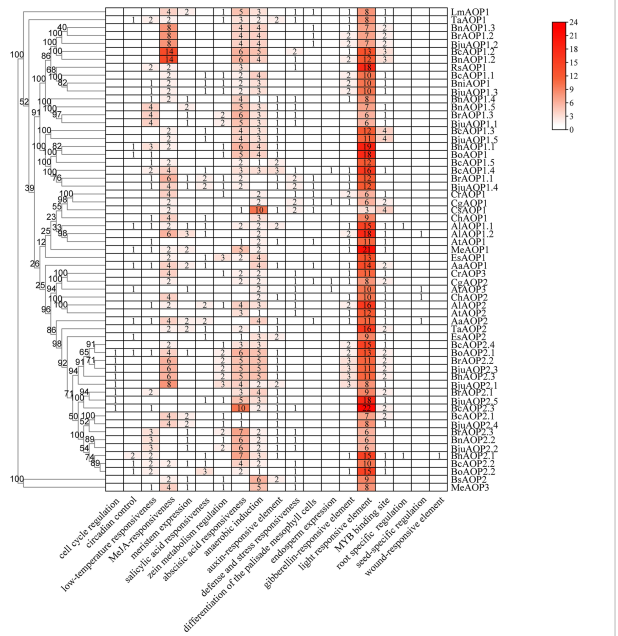


FIGURE 4
Cis-acting regulatory elements in the promoter region of *AOP* genes. The numbers and the depth of red represent the frequency of the elements that occur in the promoter region.

4 Discussion

Increasing the beneficial GSL content has been a major breeding goal for Brassica species in recent years (Ishida et al., 2014; Barba et al., 2016). The hydrolysis products of GRA such as isothiocyanates have anticancer effects in humans (Fahey et al., 2001; Fahey et al., 2002). Furthermore, enzymatic degradation of GRA to PRO inhibits iodine uptake in mammals, leading to thyroid dysfunction. The high content of SIN in TSM adds a strong mustard spiciness, which is derived from the degradation of GIB by the *AOP2* gene. Therefore, the focus of breeders' attention has been the buildup of beneficial GSL components and contents in Brassica vegetables and the reduction of harmful GSL components and contents. In *B. juncea*, Augustine and Bisht (2015) used RNA interference technology to accumulate a large amount of GRA, a beneficial GSL product, by affecting the expression of *AOP2*, indicating the significance of *AOP* genes in the breeding improvement of Brassica. The recent publication of the *B. carinata* reference genome announced the completion of the reference genome assembly for the U's Triangle Brassica Species (Song et al., 2021), which facilitates the study of the functional and evolutionary relationships of *AOP* genes. In this study, five *BjuAOP2* genes in TSM were identified and cloned. All five genes possess biological catalytic activity, but the expression and subcellular localization results are tissue-specific and differential, respectively. Therefore, functional divergence may exist in the

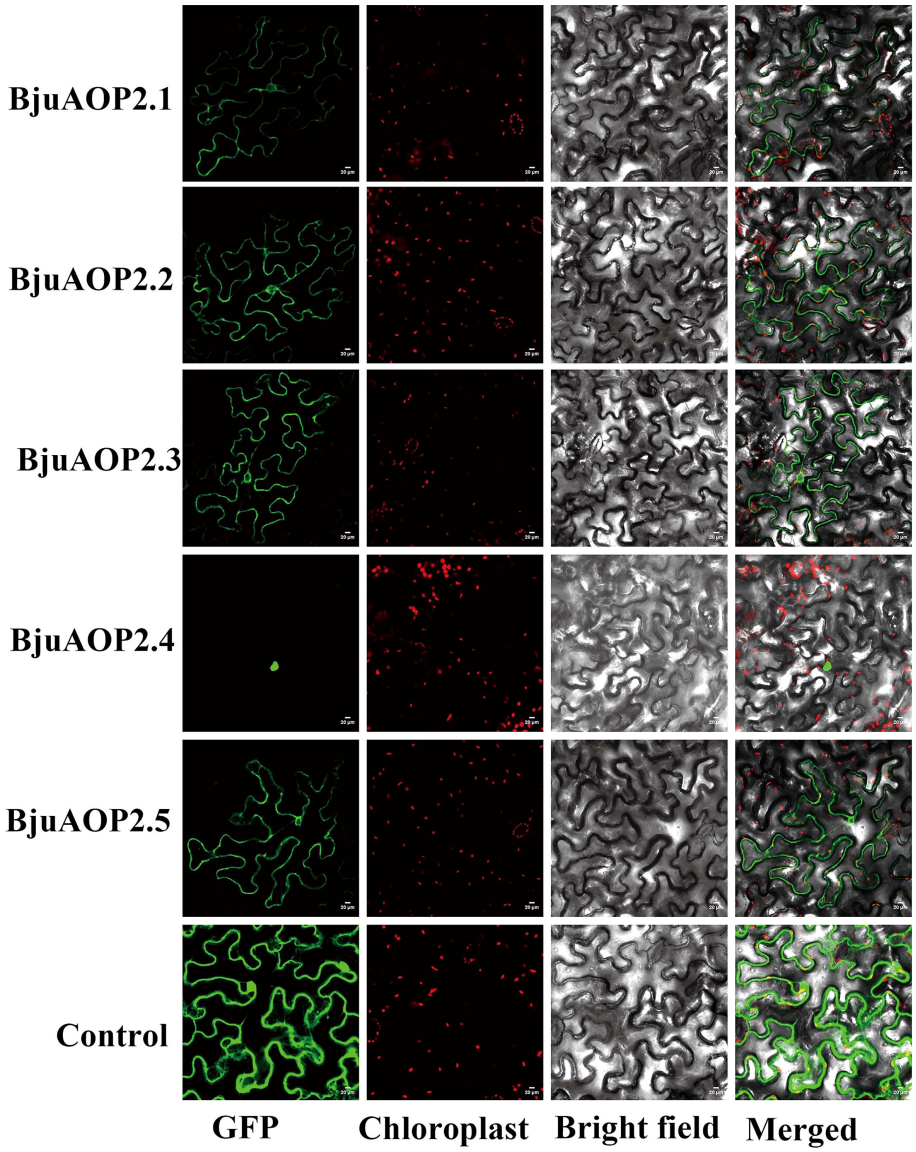


FIGURE 5
Subcellular localization of BjuAOP2 proteins in *Nicotiana benthamiana*. These genes were fused with the GFP protein driven by the 35S promoter. Green fluorescence and chloroplast autofluorescence (red) images were captured in a dark field; however, a bright field was used to get images of cell appearance. Bar=20 μm.

regulation of SIN synthesis by these five genes. Overall, the findings of this study may provide a reference for breeders to improve the TSM GSL fraction and content.

As a family with large number of members, the Cruciferae contains 338 genera and 3709 species covering many crops that have economic importance (Warwick et al., 2006). The core crops of the Cruciferae family underwent genomic triploidization events during evolution (Lysak et al., 2005). Theoretically, the number of TSM *BjuAOP2* should be three times higher (six) than that of *Arabidopsis thaliana*. However, only five *BjuAOP2* family members were retained during the evolution of TSM, indicating the loss of *BjuAOP2* genes in this process. The evolutionary analysis of cruciferous *AOP* genes in this study revealed that *AOP1* was in a separate group (Figure 2). Previous studies have shown that the *AOP* locus in cruciferous species underwent two gene duplication events. The first caused the divergence between *AOP1* and *AOP2/3* (the *AOP1* gene was the ancestor gene of the *AOP2/3* gene) while the second event led to the formation of the AOP2 and AOP3 genes (Kliebenstein et al., 2001a). Subsequently, some *AOP* genes were lost due to the formation of new species or adaptation to variable environments.

Gene duplication enables the tissue-specific expression of genes that undergo replication in response to variable environmental stimuli; in addition, the replicated genes have more diverse expression patterns relative to single genes (Li et al., 2005; Kliebenstein, 2008; Huang et al., 2022). The five *BjuAOP2* genes in this study exhibited expression divergence in different tissues and different periods of tumorous stem development, indicating varied regulation of the genes. The *BjuAOP2* gene is expressed in trace amounts in roots and in higher amounts in leaves and stems in B186

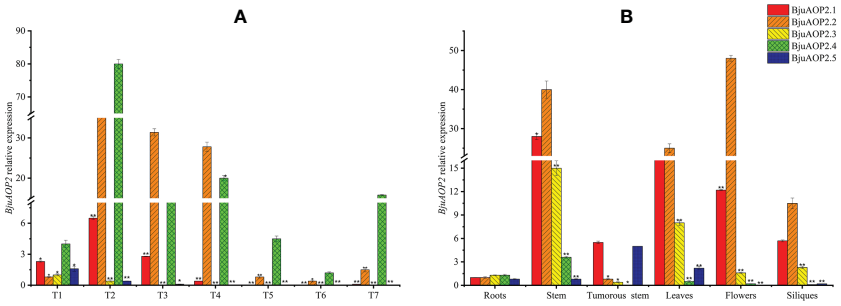


FIGURE 6
Expression pattern of *BjuAOP2* in different tissues and different developmental periods. The vertical axis indicates gene expression levels relative to *BjuAOP2.5* expression in roots **(A)** and *BjuAOP2.2* expression in 4 weeks after sowing. T1-T7 represent the seven times 4, 7, 10, 13, 16, 19, and 22 weeks after sowing **(B)**. Error bars represent the standard deviation from three biological repeats. * and ** indicate significant differences at P<0.05 and P<0.01 using ANOVA analysis followed by a Duncan test, respectively.

(Figure 6A), depicting congruency with the expression pattern of the *AOP2* gene in *Arabidopsis thaliana* (Neal et al., 2010). Subcellular localization analysis indicated differences in the expression sites of these five genes. The promoter regions of the *BjuAOP2* genes were further analyzed using PLACE software, and each of them had some specific cis-acting elements (Figure 4). These elements were associated with tissue-dependent expression, hormone, biotic and abiotic stress responses, resulting in differential expression of the five *BjuAOP2* genes in various tissues and developmental periods. Thereby leading to differences in GSL accumulation in TSM thus improving plant adaptation to the external environment.

Genes may undergo depletion, defunctionalization, functional maintenance, or functional differentiation across copies during replication (Zhang, 2003). In this research, *in vitro* experiments of prokaryotic expression demonstrated that all five genes in TSM can convert GIB to SIN, leading to abundant SIN in TSM. Therefore, all five genes have biocatalytic activity (Figure 7). The multiple-

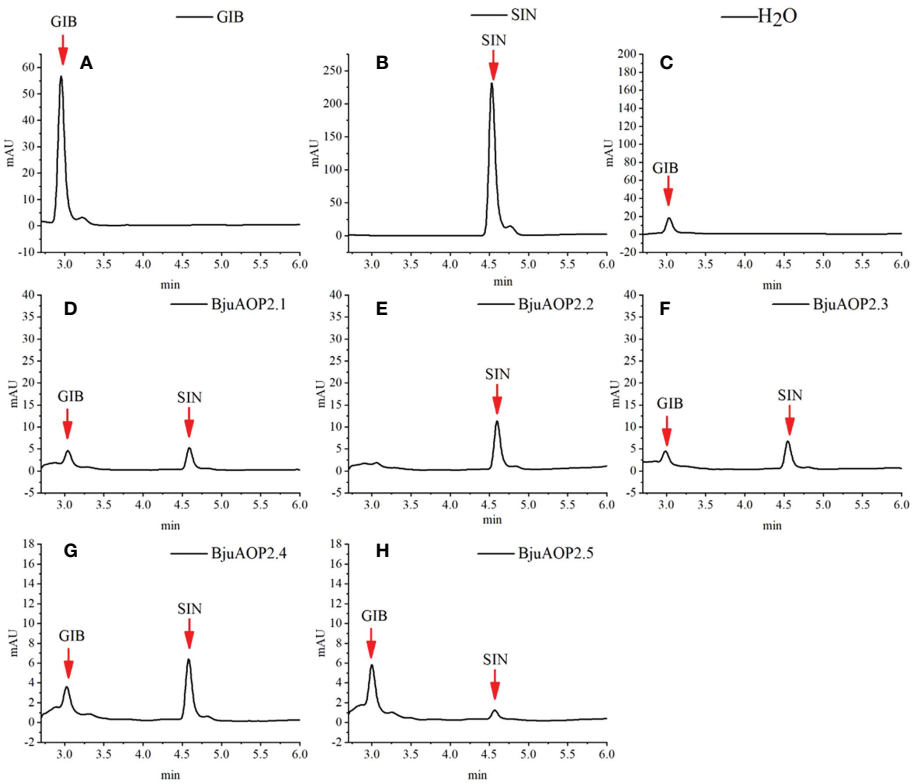


FIGURE 7
Enzymatic activity of heterologously expressed *BjuAOP2*. The purified desulfoglucosinolates extracted from *E. coli* were subjected to HPLC (monitored at 229 nm) and results were documented. **(A)** indicates desulfated GIB (glucoiberin) standard, **(B)** indicates desulfated SIN (sinigrin) standard, **(C)** indicates desulfated GIB standard treated with ddH_2O as the negative control; **(D–H)** indicate desulfated GIB standard treated with *BjuAOP2.1, BjuAOP2.2, BjuAOP2.3, BjuAOP2.4* and *BjuAOP2.5* enzymes, respectively.

sequence alignment of the five *BjuAOP2* proteins showed that they have two typically conserved structural domains at both their C- and N-terminal ends, whereas the sequence variation outside the conserved structural domain is large (Figure 1). Studies in *B. oleracea*, *B. rapa*, and *Arabidopsis thaliana* have revealed that disruption of the 2OG-FeII_Oxy structural domain leads to loss of *AOP2* gene function (Li and Quiros, 2002; Neal et al., 2010; Zhang et al., 2015). The integrity of the 2OG-FeII_Oxy structural domain may be critical for the *AOP2* protein to maintain its catalytic activity in a variable environment.

This research dealt with the elucidation of the expression characteristics of the *BjuAOP2* gene in TSM and its role in the GSL biosynthesis pathway. In addition, the evolutionary relationships of this gene were analyzed. These findings provide a guide for breeders to improve the aliphatic GSL component of TSM by traditional breeding methods or genetic modification methods.

5 Conclusion

This work identified five *BjuAOP2* genes from the TSM genome that were cloned, bioinformatically analyzed, and validated for *in vitro* activity. First, the protein sequences of the genes were analyzed. All five genes in TSM possessed two conserved structural domains, with increased variation in the variable regions, indicating functional divergence. Subsequently, the evolution of *AOP* genes in cruciferous species was analyzed. All *AOP1* genes clustered into a large independent branch, whereas *AOP2* and *AOP3* were clustered into a separate large branch. The *in vitro* activity analysis verified the biocatalytic activity of these five *BjuAOP2* genes. In addition, their expression pattern analysis suggested variation in different growth periods or different tissues. Finally, tobacco injection-based subcellular localization analysis using transient expression assay indicated the differences in the cellular localization of these five *BjuAOP2* genes.

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

Conceived and designed the experiments: JC, BC and YiL. Performed the experiments: BC, YuL, DZ, CX and ZL. Analyzed the data: JC, BC and YiL. Funding acquisition, Project administration: JC and YiL. Writing, reviewing and editing: JC, BC and YiL. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Chongqing University Innovation Research Group Funding Program (CXQT21029), the Chongqing Natural Science Foundation (cstc2019jcyj-msxmX0713), the science and technology research program of Chongqing municipal commission (KJQN202001428), the science and technology research program of Chongqing municipal commission (KJQN202001429).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1111418/full#supplementary-material

References

Augustine, R., and Bisht, N. C. (2015). Biofortification of oilseed *Brassica juncea* with the anti-cancer compound glucoraphanin by suppressing *GSL-ALK* gene family. *Sci. Rep.* 5 (1), 1–12. doi: 10.1038/srep18005

Barba, F. J., Nikmaram, N., Roohinejad, S., Khelfa, A., Zhu, Z., Koubaa, M., et al. (2016). Bioavailability of glucosinolates and their breakdown products: Impact of processing. *Front. Nutr.* 3. doi: 10.3389/fnut.2016.00024.eCollection2016

Blažević, I., Montaut, S., Burcul, F., Olsen, C. E., Burow, M., Rollin, P., et al. (2019). Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* 169, 112100. doi: 10.1016/j.phytochem.2019.112100

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1101/289660

Chen, J. Y., Ullah, C., Reichelt, M., Beran, F., Yang, Z. L., Gershenzon, J., et al. (2020). The phytopathogenic fungus sclerotinia sclerotiorum detoxifies plant glucosinolate hydrolysis products *via* an isothiocyanate hydrolase. *Nat. Commun.* 11 (1), 3090. doi: 10.1038/s41467-020-16921-2

Cheng, F., Sun, R. F., Hou, X. L., Zheng, H. K., Zhang, F. L., Zhang, Y. Y., et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat. Genet.* 48 (10), 1218–1224. doi: 10.1038/ng.3634

Chopra, R., Johnson, E. B., Emenecker, R., Cahoon, E. B., Lyons, J., Kliebenstein, D. J., et al. (2020). Identification and stacking of crucial traits required for the domestication of pennycress. *Nat. Food* 1 (1), 84–91. doi: 10.1038/s43016-019-0007-z

Clay, N. K., Adio, A. M., Denoux, C., Jander, G., and Ausubel, F. M. (2009). Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science* 323 (5910), 95–101. doi: 10.1126/science.11646

Engel, E., Martin, N., and Issanchou, S. (2006). Sensitivity to allyl isothiocyanate, dimethyl trisulfide, sinigrin, and cooked cauliflower consumption. *Appetite* 46 (3), 263–269. doi: 10.1016/j.appet.2006.01.007

Fahey, J. W., Haristoy, X., Dolan, P. M., Kensler, T. W., Scholtus, I., Stephenson, K. K., et al. (2002). Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of *Helicobacter pylori* and prevents benzo[a]pyrene-induced stomach tumors. *PNAS* 99 (11), 7610–7615. doi: 10.1073/pnas.112203099

Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56 (1), 5–51. doi: 10.1016/S0031-9422(00)00316-2

Gao, M., Li, G., Yang, B., McCombie, W. R., and Quiros, C. F. (2004). Comparative analysis of a *Brassica* BAC clone containing several major aliphatic glucosinolate genes with its corresponding *Arabidopsis* sequence. *Genome* 47 (4), 666–679. doi: 10.1139/g04-021

Grubb, C. D., and Abel, S. (2006). Glucosinolate metabolism and its control. *Trends Plant Sci.* 11 (2), 89–100. doi: 10.1016/j.tplants.2005.12.006

Grubb, C. D., Zipp, B. J., Ludwig-Müller, J., Masuno, M. N., Molinski, T. F., and Abel, S. (2004). Arabidopsis glucosyltransferase *UGT74B1* functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* 40 (6), 893–908. doi: 10.1111/j.1365-313X.2004.02261.x

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57 (1), 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Hansen, B. G., Kerwin, R. E., Ober, J. A., Lambrix, V. M., Mitchell-Olds, T., Gershenzon, J., et al. (2008). A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. *Plant Physiol.* 148 (4), 2096–2108. doi: 10.1104/pp.108.129981

Harun, S., Abdullah-Zawawi, M. R., Goh, H. H., and Mohamed-Hussein, Z. A. (2020). A comprehensive gene inventory for glucosinolate biosynthetic pathway in *Arabidopsis thaliana*. *J. Agric. Food Chem.* 68 (28), 7281–7297. doi: 10.1021/acs.jafc.0c01916

He, J. H., Chen, H., and Schnitzler, W. H. (2002). Glucosinolate composition and contents in brassica vegetables. *Scientia Agric. Sin.* 35 (2), 192–197. doi: 10.1006/jfls.2001.0409

Huang, Y., Chen, J., Dong, C., Sosa, D., Xia, S., Ouyang, Y., et al. (2022). Species-specific partial gene duplication in *Arabidopsis thaliana* evolved novel phenotypic effects on morphological traits under strong positive selection. *Plant Cell* 34 (2), 802–817. doi: 10.1101/2021.04.05.438504

Ishida, M., Hara, M., Fukino, N., Kakizaki, T., and Morimitsu, Y. (2014). Glucosinolate metabolism, functionality and breeding for the improvement of brassicaceae vegetables. *Breed Sci.* 64 (1), 48–59. doi: 10.1270/jsbbs.64.48

Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinf.* 11, 431. doi: 10.1186/1471-2105-11-431

Kliebenstein, D. J. (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PloS One* 3 (3), e1838. doi: 10.1371/journal.pone.0001838

Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., et al. (2001a). Genetic control of natural variation in arabidopsis glucosinolate accumulation. *Plant Physiol.* 126 (2), 811–825. doi: 10.1104/pp.126.2.811

Kliebenstein, D. J., Lambrix, V. M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001b). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell* 13 (3), 681–693. doi: 10.1105/tpc.13.3.681

Kroymann, J., Donnerhacke, S., Schnabelrauch, D., and Mitchell-Olds, T. (2003). Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *PNAS* 100 (suppl-2), 14587–14592. doi: 10.1073/pnas.1734046100

Kroymann, J., Textor, S., Tokuhisa, J. G., Falk, K. L., Bartram, S., Gershenzon, J., et al. (2001). A gene controlling variation in arabidopsis glucosinolate composition is part of the mcthionine chain elongation pathway. *Plant Physiol.* 127 (3), 1077–1088. doi: 10.1104/pp.010416

Li, B. H., Gaudinier, A., Tang, M., Taylor-Teeples, M., Nham, N. T., Ghaffari, C., et al. (2014). Promoter-based integration in plant defense regulation. *Plant Physiol.* 166 (4), 1803–1820. doi: 10.1104/pp.114.248716

Li, W. H., Jing, Y., and Xun, G. (2005). Expression divergence between duplicate genes. *Trends Genet.* 21 (11), 602–607. doi: 10.1016/j.tig.2005.08.006

Li, G., and Quiros, C. F. (2002). Genetic analysis, expression and molecular characterizatien of BoGSL-ELONG, a major gene involved in the aliphatic glucosinolate pathway of brassica species. *Genetics* 162 (4), 1937–1943. doi: 10.1017/S0016672302005906

Li, G., and Quiros, C. F. (2003). In planta side-chain glucosinolate modification in arabidopsis by introduction of dioxygenase brassica homolog BoGSL-ALK. *Theor. Appl. Genet.* 106 (6), 1116–1121. doi: 10.1007/s00122-002-1161-4

Li, B. H., Tang, M., Nelson, A., Caligagan, H., Zhou, X., Clark-Wiest, C., et al. (2018). Network-guided discovery of extensive epistasis between transcription factors involved in aliphatic glucosinolate biosynthesis. *Plant Cell* 30 (1), 178–195. doi: 10.1105/tpc.17.00805

Li, Y., Wang, X. Y., Wang, Y. H., Meng, Q. F., Sun, J., and Wang, B. L. (2011). Studies on composition and contents of glucosinolates in different tuber mustard varieties. *Acta Hortic. Sin.* 38 (7), 1356–1364. doi: 10.16420/j.issn.0513-353x.2011.07.001

Liu, Z., Hirani, A. H., Mcvetty, P. B., Daayf, F., Quiros, C. F., and Li, G. (2012). Reducing progoitrin and enriching glucoraphanin in *Braasica napus* seeds through silencing of the GSL-ALK gene family. *Plant Mol. Biol.* 79 (1), 179–189. doi: 10.1007/s11103-012-9905-2

Liu, Z., Li, N., Yu, T., Wang, Z., Wang, J., Ren, J., et al. (2022). The brassicaceae genome resource (TBGR): A comprehensive genome platform for brassicaceae plants. *Plant Physiol.* 190 (1), 226–237. doi: 10.1093/plphys/kiac266

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5 (1), 1–11. doi: 10.1038/ncomms4930

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 (4), 402–408. doi: 10.1006/meth.2001.1262

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe *Brassiceae*. *Genome Res.* 15 (4), 516–525. doi: 10.1101/gr.3531105

Malhotra, B., and Bisht, N. C. (2020). Glucosinolates: regulation of biosynthesis and hydrolysis. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.620965

Mazumder, A., Dwivedi, A., and Du Plessis, J. (2016). Sinigrin and its therapeutic benefits. *Molecules* 21 (4), 416. doi: 10.3390/molecules21040416

Mikkelsen, M. D., Naur, P., and Halkier, B. A. (2004). *Arabidopsis* mutants in the c-s lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* 37 (5), 770–777. doi: 10.1111/j.1365-313x.2004.02002.x

Mitreiter, S., and Gigolashvili, T. (2021). Regulation of glucosinolate biosynthesis. *J. Exp. Bot.* 72 (1), 70–91. doi: 10.1093/jxb/eraa479

Neal, C. S., Fredericks, D. P., Griffiths, C. A., and Neale, A. D. (2010). The characterisation of *AOP2*: a gene associated with the biosynthesis of aliphatic alkenyl glucosinolates in *Arabidopsis thaliana*. *BMC Plant Biol.* 10 (1), 1–16. doi: 10.1186/1471-2229-10-170

Nguyen, V. P., Stewart, J., Lopez, M., Ioannou, I., and Allais, F. (2020). Glucosinolates: natural occurrence, biosynthesis, accessibility, isolation, structures, and biological activities. *Molecules* 25 (19), 4537. doi: 10.3390/molecules25194537

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. *Cell Res.* 19 (2), 271–273. doi: 10.1038/cr.2009.6

Sønderby, I. E., Geu-Flores, F., and Halkier, B. A. (2010). Biosynthesis of glucosinolates-gene discovery and beyond. *Trends Plant Sci.* 15 (5), 283–290. doi: 10.1016/j.tplants.2010.02.005

Song, X. M., Wei, Y. P., Xiao, D., Gong, K., Sun, P. C., Ren, Y. M., et al. (2021). *Brassica carinata* genome characterization clarifies u's triangle model of evolution and polyploidy in *Brassica*. *Plant Physiol.* 186 (1), 388–406. doi: 10.1093/plphys/kiab048

Soundararajan, P., and Kim, J. S. (2018). Anti-carcinogenic glucosinolates in cruciferous vegetables and their antagonistic effects on prevention of cancers. *Molecules* 23 (11), 2983. doi: 10.3390/molecules23112983

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi: 10.1093/molbev/mst197

Textor, S., Kraker de, J. W., Hause, B., Gershenzon, J., and Tokuhisa, J. G. (2007). *MAM3* catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiol.* 144 (1), 60–71. doi: 10.1104/pp.106.091579

Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., et al. (2011). Glucosinolate biosynthetic genes in *Brassica rapa*. *Gene* 487 (2), 135–142. doi: 10.1016/j.gene.2011.07.021

Warwick, S. I., Francis, A., and Al-Shehbaz, I. A. (2006). *Brassicaceae*: Species checklist and database on CD-rom. *Plant Systematics Evol.* 259 (2), 249–258. doi: 10.1007/s00606-006-0422-0

Wittstock, U., and Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis* era. *Trends Plant Sci.* 7 (6), 263–270. doi: 10.1016/S1360-1385(02)02273-2

Wu, S. H., Lei, J. J., Chen, G. J., Chen, H. C., Cao, B. H., and Chen, C. M. (2017). *De novo* transcriptome assembly of Chinese kale and global expression analysis of genes involved in glucosinolate metabolism in multiple tissues. *Front. Plant Sci.* 8 (92). doi: 10.3389/fpls.2017.00092

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18 (6), 292–298. doi: 10.1016/S0169-5347(03)00033-8

Zhang, J. F., Liu, Z. Y., Liang, J. L., Wu, J., Cheng, F., and Wang, X. W. (2015). Three genes encoding AOP2, a protein involved in aliphatic glucosinolate biosynthesis, are differentially expressed in. *Brassica rapa. J. Exp. Bot.* 66 (20), 6205–6218. doi: 10.1093/jxb/erv331

Zheng, H., Wang, Y., Li, X., Huang, W., Miao, H., Li, H., et al. (2022). A novel putative 2-oxoglutarate-dependent dioxygenase gene (*BoaAOP-like*) regulates aliphatic glucosinolate biosynthesis in Chinese kale. *Scientia Hortic.* 297, 110921. doi: 10.1016/j.scienta.2022.110921

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/fpls.2023.1135684

OPEN ACCESS

EDITED BY
Yoonkang Hur,
Chungnam National University,
Republic of Korea

REVIEWED BY
Zhansheng Li,
Insititute of Vegetables and Flowers,
Chinese Academy of Agricultural Sciences (CAAS), China
Jianwei Gao,
Shandong Academy of Agricultural Sciences, China

*CORRESPONDENCE
Xiaonan Li
gracesleexn@163.com
Hong Lang
langhong@jlnku.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 01 January 2023
ACCEPTED 15 February 2023
PUBLISHED 24 February 2023

CITATION
Jiang M, Zhang Y, Yang X, Li X and Lang H (2023) *Brassica rapa* orphan gene *BR1* delays flowering time in *Arabidopsis*. *Front. Plant Sci.* 14:1135684. doi: 10.3389/fpls.2023.1135684

Brassica rapa orphan gene *BR1* delays flowering time in *Arabidopsis*

Mingliang Jiang[1†], Yuting Zhang[2†], Xiaolong Yang[3], Xiaonan Li[2*] and Hong Lang[1*]

[1]School of Agriculture, Jilin Agricultural Science and Technology College, Jilin, China, [2]College of Horticulture, Shenyang Agricultural University, Shenyang, China, [3]College of Horticulture, South China Agricultural University, Guangzhou, China

Orphan genes are essential to the emergence of species-specific traits and the process of evolution, lacking sequence similarity to any other identified genes. As they lack recognizable domains or functional motifs, however, efforts to characterize these orphan genes are often difficult. Flowering is a key trait in *Brassica rapa*, as premature bolting can have a pronounced adverse impact on plant quality and yield. Bolting resistance-related orphan genes, however, have yet to be characterized. In this study, an orphan gene designated *BOLTING RESISTANCE 1* (*BR1*) was identified and found through gene structural variation analyses to be more highly conserved in Chinese cabbage than in other available accessions. The expression of *BR1* was increased in bolting resistant Chinese cabbage and decreased in bolting non-resistant type, and the expression of some mark genes were consist with bolting resistance phenotype. *BR1* is primarily expressed in leaves at the vegetative growth stage, and the highest *BR1* expression levels during the flowering stage were observed in the flower buds and silique as compared to other tissue types. The overexpression of *BR1* in *Arabidopsis* was associated with enhanced bolting resistance under long day (LD) conditions, with these transgenic plants exhibiting significant decreases in stem height, rosette radius, and chlorophyll content. Transcriptomic sequencing of WT and *BR1*OE plants showed the association of *BR1* with other bolting resistance genes. Transcriptomic sequencing and qPCR revealed that six flowering integrator genes and one chlorophyll biosynthesis-related gene were downregulated following *BR1* overexpression. Six key genes in photoperiodic flowering pathway exhibited downward expression trends in *BR1*OE plants, while the expression of floral repressor *AtFLC* gene was upregulated. The transcripts of these key genes were consistent with observed phenotypes in *BR1*OE plants, and the results indicated that *BR1* may function through vernalization and photoperiodic pathway. Instead, the protein encoded by *BR1* gene was subsequently found to localize to the nucleus. Taken together, we first propose that orphan gene *BR1* functions as a novel regulator of flowering time, and these results suggested that *BR1* may represent a promising candidate gene to support the selective breeding of Chinese cabbage cultivars with enhanced bolting resistance.

KEYWORDS

Brassica rapa*, orphan gene, *BR1*, bolting resistance, *Arabidopsis

1 Introduction

Orphan genes (*OGs*) are species- or lineage-specific genes that lack sequence similarity with other known genes expressed by other species (Jiang et al., 2022), often arising as a result of rapid evolutionary activity (Cai et al., 2006). To date, *OGs* have been reported in many genomic sequencing analyses of species including *Aegiceras corniculatum* (Ma et al., 2021), *Arabidopsis thaliana* (Donoghue et al., 2011), *Brassica rapa* (Jiang et al., 2018), eight Cucurbitaceae species (Ma et al., 2022), *Oryza sativa* (Cui et al., 2015), and *Vigna unguiculata* (Li et al., 2019). As *OGs* do not contain recognizable domains, functional motifs, or folding patterns, their functions are often unclear such that detailed functional characterization is ultimately necessary. Several studies have successfully demonstrated the diverse roles played by specific *OGs* as mediators of metabolite synthesis (Li et al., 2009; Li and Wurtele, 2015; Li et al., 2015b; O'Conner et al., 2018; Jiang et al., 2020b; Fang et al., 2021; Shen et al., 2021; Tanvir et al., 2022b), biotic stresses response (Jiang et al., 2018; Qi et al., 2019; Brennan et al., 2020; Jiang et al., 2020a; Wang et al., 2021; Moon et al., 2022; Tanvir et al., 2022a), abiotic stresses response (Yadeta et al., 2014; Li et al., 2019; Ma et al., 2020), species-specific traits, or the regulation of growth and development (Chen et al., 2017; Ni et al., 2017; Wang et al., 2017; Zhao et al., 2018; Dossa et al., 2021). These characteristics make *OGs* important targets for plant breeding efforts aimed at enhancing specific traits of interest to improve plant stress resistance and quality. However, the ability of specific *OGs* to regulate the timing of flowering has largely been overlooked to date.

Flowering timing is an agronomically important trait that can determine reproductive success and shape crop yields. This timing is thus regulated by a complex network of signaling proteins and processes that are responsive to external stimuli and developmental cues (Wang, 2014; Blümel et al., 2015). Five genetically defined pathways have been identified to date as important regulators of floral transition, including the age, photoperiod, hormone, autonomous, and vernalization pathways (Srikanth and Schmid, 2011; Li et al., 2015a; Freytes et al., 2021). These pathways integrate diverse signaling inputs associated with flowering and ultimately regulate the expression of important genes that govern flowering timing including *APETALA1* (*AP1*, *AT1G69120*), *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*, *AT2G45660*), *FLOWERING LOCUS T* (*FT*, *AT1G65480*), and the plant-specific transcription factor (TF) *LEAFY* (*LFY*, *AT5G61850*) (Wang, 2014; Bao et al., 2020). The autonomous pathway can promote flowering in a manner that is independent of the length of the day through the suppression of *FLOWERING LOCUS C* (*FLC*, *AT5G10140*), which is a central repressor of flowering and vernalization (Simpson and Dean, 2002; Jung and Müller, 2009; Cheng et al., 2017). FLC is an important inhibitor of flowering activity that can suppress shoot apical meristem (SAM) floral transition-related TF expression (Cho et al., 2017). In the photoperiod pathway, daily patterns of expression for the floral regulators *CONSTANS* (*CO*, *AT5G15840*) and *FT* are regulated by a range of positive and negative factors that can influence protein-protein interactions, chromatin structural characteristics, protein stability, and transcriptional activity (Johansson and Staiger, 2015). Circadian rhythms are closely linked to the photoperiod-mediated control of the transition from vegetative to reproductive plant growth (Creux and Harmer, 2019). Gibberellin (GA) pathway signaling serves as a major hormonal mechanism that regulates flowering activity, although other hormones including jasmonate, brassinosteroid, abscisic acid, cytokinins, and ethylene also play regulatory roles in this context (Izawa, 2021). The *miR156-SPL* (*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE*) and *miR172-AP2* (*APETALA2*, *AT4G36920*) modules have been suggested as key regulatory hubs involved in the age pathway that facilitate plant flowering under non-inductive conditions (Wang, 2014; Kinoshita and Richter, 2020). While these results provide important insight into the mechanisms that control the timing of flowering, the mechanistic links among these pathways and the regulatory crosstalk between them have yet to be characterized in detail.

Heading is among the most important agronomic traits for *B. rapa* ssp. *pekinensis* (Chinese cabbage) or *Brassica oleracea* var. *capitata* (cabbage) and has been the subject of extensive research interest (Zhang et al., 2022). Premature bolting can have a severe adverse impact on Chinese cabbage or cabbage yields and quality, restricting the geographic distribution and planting season for this economically important species (Su et al., 2018). Accordingly, there is a pressing need to breed novel cultivars with enhanced bolting resistance. The histone H4 protein encoded by *BrHIS4.A04* (*Bra035673*) has previously been shown to attenuate photoperiod-related flowering gene expression under drought conditions in Chinese cabbage plants *via* signaling through the ABA pathway, thus preventing premature bolting (Xin et al., 2020). *SET DOMAIN GROUP 8* (*BrSDG8*, *BraA07g040740.3C*) serves as an additional regulator of early bolting in *B. rapa* ssp. *pekinensis*, with *FLC* H3K6 methylation activity increasing when the function of this gene is disrupted (Fu et al., 2020). *BrFLC5* (*Bra022771*) was also previously found to be expressed at lower levels than two other *BrFLC* genes, supporting efforts to breed *B. rapa* plants resistant to premature bolting (Xi et al., 2018). Cabbage *BoFLC4-1* played a similar role to *Arabidopsis FLC* in regulating flowering time (Lin et al., 2005). The intron I 215-bp indel of *BoFLC2* influenced the flowering time of cabbage, which might offer critical information to promote the study of epigenetic gene silencing processes in flowering-related genes (Li et al., 2022). Study also showed that *BoFLC1*, *BoFLC3*, and *BoFLC5* were within the confidence intervals of known flowering time quantitative trait loci (QTL) (Razi et al., 2008). These prior results thus offer valuable insights that can be leveraged to better facilitate the genetic control of the bolting and flowering processes.

In this study, the *OG* designated *BOLTING RESISTANCE 1* (*BR1*) was the target of functional characterization efforts exploring its relationship with the timing of flowering. Structural genotypic variations and sequence characteristics in the *BR1* gene region were analyzed in 524 *B. rapa* accessions, and its expressions in bolting resistant or bolting non-resistant Chinese cabbage were detected. Representative mark genes were determined in bolting resistant type inbred lines. And *BR1* expression patterns were examined over the course of Chinese cabbage development in different tissue

compartments. The impact of *BR1* overexpression on flowering time was additionally assessed in *Arabidopsis thaliana*, while transcriptomic sequencing was used to explore the mechanistic basis for the ability of *BR1* to delay *A. thaliana* flowering time. Subcellular localization of the BR1 protein was additionally analyzed. Together, these analyses identified the *B. rapa* ssp. *pekinensis BR1* gene as a key regulator of delayed flowering time in *A. thaliana*.

2 Materials and methods

2.1 Plant material and cultivation

B. rapa 'Chiifu' cultivar, Chinese cabbage inbred lines, *A. thaliana* ecotype 'Columbia-0' (Col-0), and the transgenic *A. thaliana* lines were cultivated as in our prior studies (Jiang et al., 2018; Jiang et al., 2020b). *Nicotiana benthamiana* was cultivated as in prior study (Zhan et al., 2022).

2.2 *BR1* sequence analyses

The CD-Search tool was used to search the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for conserved domains in BR1. The SignalP 5.0 server (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) was used to identify predicted signal peptide sequences in BR1. The PROSITE database (https://prosite.expasy.org/) was used to predict motifs in BR1, while TF predictions were made with the Plant Transcription Factor Database (PlantTFDB v5.0, http://planttfdb.gao-lab.org/). Genotypic analyses of structural variations in the *BR1* gene region were performed by comparing 524 *B. rapa* accessions with the Polymorph tool using the Brassicaceae Database (BRAD, http://brassicadb.cn/) as reported previously (Cai et al., 2021), with these 524 different *B. rapa* accessions being derived from a separate report (Cheng et al., 2016; Su et al., 2018; Cai et al., 2021). Ten orphan genes analyzed in this study were identified in our previous study (Jiang et al., 2018). The *B. rapa* genome version 3.0 was used for structural variation analyses.

2.3 Analyses of *BR1* expression profiles in Chinese cabbage

Quantitative real-time PCR (qPCR) analyses were performed as detailed previously (Jiang et al., 2018), using primers compiled in Supplementary Table S1. Sampling of plants at the seedling and flowering stages was performed as published previously (Jiang et al., 2018), while sampling at the rosette and heading stages was performed at the 6th and 8th weeks, respectively. When rosette stage sampling was performed, three leaves were collected from each of nine individual 'Chiifu' plants (three biological replicates, three plants per replicate). With the oldest leaf numbered as leaf one, samples were collected from different positions (top, middle, bottom) on three different leaves (outer leaf, first leaf; middle leaf, 10th leaf; inner leaf, 20th leaf), designated as the RL1, RL2, and RL3 from the outer to the inner leaves. In the heading stage, three leaves (outer leaf, first leaf; middle leaf, 20th leaf; inner leaf, 40th leaf) were similarly collected from each of nine 'Chiifu' plants, with these samples being respectively designated as HL1, HL2, and HL3. For *BR1* expression in bolting resistant or bolting non-resistant Chinese cabbage, ten inbred lines from heading stage were selected from our laboratory, the top point of short stem (GP) and the top point of inner leaf (TP) were sampled, TP was used as control. Three biological replicates with three plants of different lines per replicate were sampled. After collection, samples were snap-frozen with liquid nitrogen and stored at −80°C for subsequent RNA isolation. qPCR primers for the expression analyses of representative mark genes were listed in Supplementary Table S1.

2.4 Establishment and analysis of *BR1*-overexpressing transgenic *Arabidopsis* plants

Vector constructs, *Arabidopsis* transformation, and selection were all performed as detailed in our prior report (Jiang et al., 2020b). Vector construction was performed using primer pairs compiled in Supplementary Table S1. Phenotypic analyses of transgenic *Arabidopsis* plants were performed as detailed in our prior study (Jiang et al., 2020b).

2.5 Transcriptomic sequencing and validation

Three biological replicate samples were collected from the aerial portions of WT and *BR1*OE mutant plants 25 days post-planting. After snap freezing using liquid nitrogen, these samples were stored at −80°C. RNA extraction was performed as detailed previously (Jiang et al., 2018), and 1% agarose gel electrophoresis was used to detect any RNA contamination or degradation while a NanoPhotometer® instrument (IMPLEN, CA, USA) was used to confirm RNA purity. A Qubit® RNA Assay Kit and a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) were used to quantify the RNA concentrations in individual samples, while an RNA Nano 6000 Assay Kit and a Bioanalyzer 2100 instrument (Agilent Technologies, CA, USA) were used to confirm RNA integrity. Sequencing libraries were prepared from 1 μg of RNA per sample with a NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) based on provided directions. Library sequencing was then performed with an Illumina Hiseq platform to generate 125 bp/150 bp paired-end reads.

Initial data were filtered with Fastp (v0.19.3) to remove adapter-containing reads, reads containing > 10% N bases, and reads with > 50% low-quality (Q ≤ 20) bases. The clean reads were then compared to the *Arabidopsis* TAIR10 genome which was downloaded from The *Arabidopsis* Information Resource (TAIR) (https://www.arabidopsis.org/) using HISAT (v2.1.0). New gene

predictions were made using StringTie (v1.3.4d), while gene alignment was calculated with FeatureCounts (v1.6.2), and fragments per kilobases of exons per million mapped reads (FPKM) expression values were then calculated for all transcripts. Differentially expressed genes (DEGs) were identified using DESeq2 (v1.22.1) based on Benjamini & Hochberg-corrected *p*-values, a | log2Fold Change| ≥ 1, and a false discovery rate (FDR) < 0.05. Hypergeometric tests were used for Gene Ontology (GO) term and KEGG pathway enrichment analyses. Gene expression was validated using primer pairs listed in Supplementary Table S1.

2.6 Subcellular localization analyses

The *BR1* coding sequence was cloned into the pCAM35-GFP vector without the corresponding stop codons using the KpnI and BamHI cleavage sites. pCAM35::BR1::GFP expression vector construction and *N. benthamiana* epidermal cell infection was performed as reported previously (Zhang et al., 2021). Nuclei were visualized through the co-expression of a mCherry-labeled nuclear marker. All experiments were independently repeated in triplicate, and a Leica confocal microscope (SP8, Germany) was used to visualize cells at 48 h following agro-infiltration. Primer pairs used in vector construction are compiled in Supplementary Table S1.

2.7 Statistical analysis

SPSS 19.0 was used to compare data through Student's *t*-tests or one-way ANOVAs with Duncan's multiple range test as appropriate.

3 Results

3.1 Screening of potential bolting resistance orphan genes

Genotypic analyses of structural variations of ten orphan genes were randomly selected from our prior study (Jiang et al., 2018) (Supplementary Table S2), which was used to screen potential bolting resistance orphan genes. BRAD was used to perform genotypic analyses of structural variations in this *BR1* gene region across 524 *B. rapa* accessions. There were no any structural variations in four orphan genes (*BraA01g024790.3C*, *BraA02g027570.3C*, *BraA03g056750.3C*, and *BraA09g030600.3C*). The highest average sequence conservation rates was found in the gene region of *BR1* (*BraA10g003580.3C*), including just four single nucleotide polymorphisms (SNPs) in this region (A10_1866376, A10_1866412, A10_1866527, and A10_1866663) (Figures 1A-D), which accounting for ~80% in Chinese cabbage relative to ~46% in other accessions (Figure 1E). The A10_1866412 SNP exhibited the highest sequence conservation ratio (~98%) in Chinese cabbage, while the average variation ratio in Chinese cabbage was ~20%, with this value being lower than in other accessions (~54%). The A10_1866376 SNP exhibited the highest variation ratio (~76%) in other accessions. Average *BR1* sequence conservation rates were approximately 77%, 75%, and 88% in Chinese cabbage spring, summer, and autumn ecotypes, respectively (Figure 1F). No variations in the A10_1866412 SNP were observed among these three Chinese cabbage ecotypes, and the average percentage of variation in these three respective ecotypes was approximately 23%, 25%, and 12%. *BR1* located on chromosome A10 at positions 1866284 – 1866805, and this gene contains no introns and consists of 522 bases that encode a protein 173 amino acids in length. The BR1 protein does not contain any known signal peptides, motifs, or conserved domains, and it was not identified as a TF in subsequent analyses. As such, the *BR1* gene is more conserved in Chinese cabbage than in other accessions, potentially owing to domestication and associated selection for leafy head development and bolting resistance in this economically important species.

3.2 *BR1* expression patterns in Chinese cabbage

qPCR was used to analyze *BR1* expression patterns in an effort to explore its potential functional roles during different stages of *B. rapa* development. Significant increases in *BR1* expression were observed in both roots and leaves in the seedling stage, with peak expression in the leaves (Figure 2A). During the rosette and heading stages, *BR1* expression levels rose significantly in the middle (RL2 and HL2) and inner leaves (RL3 and HL3) relative to the outer leaves (RL1 and HL1) (Figures 2B, C). During the flowering stage, higher *BR1* expression levels were detected in the flower buds and silique relative to other analyzed tissues (Figure 2D). These results provide a basis for the further exploration of *BR1* as an *OG* associated with the regulation of different stages of vegetative and reproductive growth in Chinese cabbage plants. Next, the expression levels of *BR1* were detected within bolting resistant (BR type) and bolting non-resistant (BN type) Chinese cabbage to explore its functions. Study indicated that Chinese cabbage with the rounded apices of its short stem belonging to BR type, and BN type possessed the pointed apices (Mero and Honma, 1984). Surprisingly, *BR1* possessed increase trends in five lines from BR type, but showed suppressed trends in BN type (Figure 3A). Meanwhile, the expression patterns of mark genes were analyzed to confirm the function of controlling bolting resistance in Chinese cabbage. Mark genes including *BrFLCs*, *BrFTs*, *BrSOC1s*, and *BrLFYs*. BR type inbred lines 'BR-98' was selected for further analysis. As expected, four *BrFLCs* showed increased trends, while *BrFTs*, *BrSOC1s*, and *BrLFY2* were displayed decreased trends (Figure 3B). The expression of *BrFT3* and *BrLFY1* was not detected. These expression patterns correspond to the bolting resistance phenotype of inbred line 'BR-98'. Such results suggested that *BR1* may directly or indirectly involve in bolting resistance of Chinese cabbage.

10.3389/fpls.2023.1135684

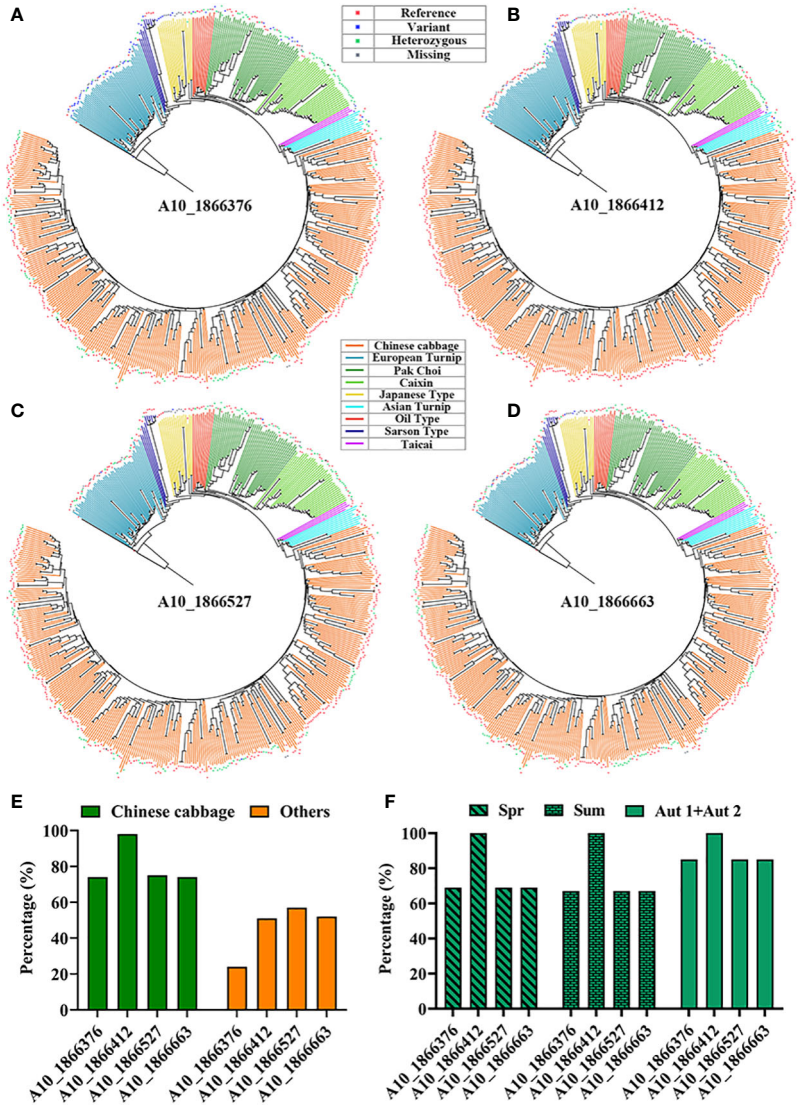


FIGURE 1
Genotypic analyses of SNPs in the *BR1* gene region. **(A-D)** Different SNP types identified across 524 *B rapa* accessions in the *BR1* gene region. If a particular accession was consistent with the reference genome, this is denoted by a reference designation, whereas the variant designation was used to indicate an accession with a genotype divergent from the reference genome. Deletion events were designated by 'Missing', and heterozygous typing was represented by 'Heterozygous'. **(E)** The percentage of sequence conservation in Chinese cabbage and other accessions. **(F)** Percentage of sequence conservation in different Chinese cabbage ecotypes, with Spr and Sum respectively corresponding to the spring and summer ecotypes, while the autumn ecotypes are denoted by Aut1+Aut2.

3.3 *BR1* overexpression regulates bolting resistance in *Arabidopsis*

To better understand the functional roles played by *BR1*, *A. thaliana* plants overexpressing this gene were prepared. After floral dip transformation, the DsRed marker was used to select T2 homozygous seeds from different self-pollinated T1 transgenic seed lines, and T3 plants were then planted to assess phenotypes following the harvesting of seeds from T2 plants. Phenotypes were compared between the wild-type (WT) and *BR1* overexpressing plants (*BR1*OE) in the vegetative and reproductive phases of growth under LD conditions (16 h of light/8 h of dark). These analyses revealed that *BR1* overexpression strongly delayed *Arabidopsis* floral transition (Figure 4A), with *BR1OE* plants exhibiting a 28% delay in flowering time relative to WT controls (Figure 4B). Rosette radius was also reduced by ~26% in these *BR1*OE mutant plants (Figure 4C), suggesting that *BR1* may regulate leaf elongation under standard growth conditions. *BR1OE* plants also exhibited a decrease in final stem height relative to WT controls (Figure 4D), whereas silique length and numbers of seeds per silique were unchanged (Figures 4E, F). Notably, *BR1*OE plants exhibited reductions in chlorophyll content relative to WT plants (Figures 4G-I). Accordingly, the overexpression of *BR1* enhances *Arabidopsis* bolting resistance.

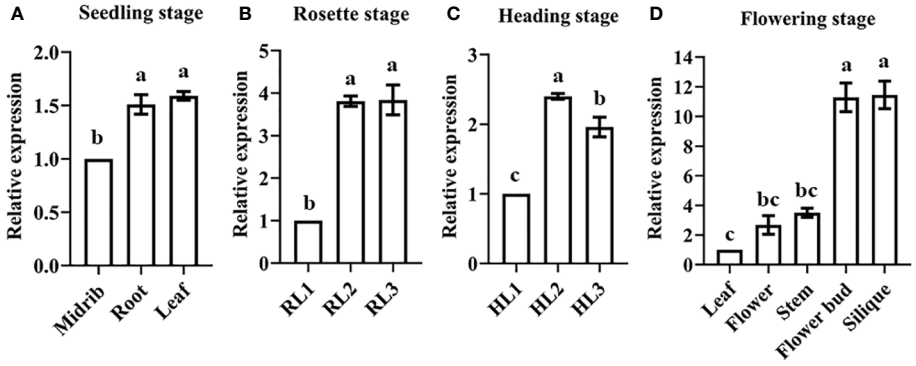


FIGURE 2
BR1 expression patterns during different developmental stages in Chinese cabbage. *BR1* expression was assessed in the **(A)** seedling, **(B)** rosette, **(C)** heading, and **(D)** flowering stages. Data are means ± SE of three independent measurements. Statistically distinct groups are marked with black letters (one-way ANOVA, $p < 0.05$).

3.4 Analysis of transcriptomic sequencing data and qPCR validation

Transcriptomic sequencing was conducted to explore the molecular mechanisms underlying the effects of *BR1* on delayed flowering time in *A. thaliana* by preparing cDNA libraries from the leaves of WT and *BR1OE* plants. All samples exhibited Q20 values > 97% and Q30 values > 92% (Supplementary Table S3). Raw and clean read numbers for individual samples respectively ranged from 43,611,780 - 48,073,534 and 42,257,488 - 46,814,082, consistent with highly reliable transcriptomic detection results.

Relative to WT plants, the *BR1*OE mutants exhibited significantly delayed flowering time. To better understand the transcriptomic changes underlying this phenotype, FPKM values for individual unigenes were compared between these two *A. thaliana* varieties to identify genes that were differentially expressed. In total, 254 DEGs were identified of which 73 and 181 were respectively up- and downregulated (28.74% and 71.26%, respectively) (Figure 5A, Supplementary Table S4). GO enrichment analyses of these DEGs identified 44 significantly enriched GO terms (Figure 5B, Supplementary Table S5). These included enriched biological processes (including cellular process, metabolic process, biological regulation, developmental process, regulation of biological process, multicellular organismal progress, signaling, reproductive process, reproduction, growth, and rhythmic process), cellular component (including cell, cell part, organelle, membrane, membrane part, and extracellular region), and molecular function (including binding, catalytic activity, transcription regulator activity, transporter activity, structural

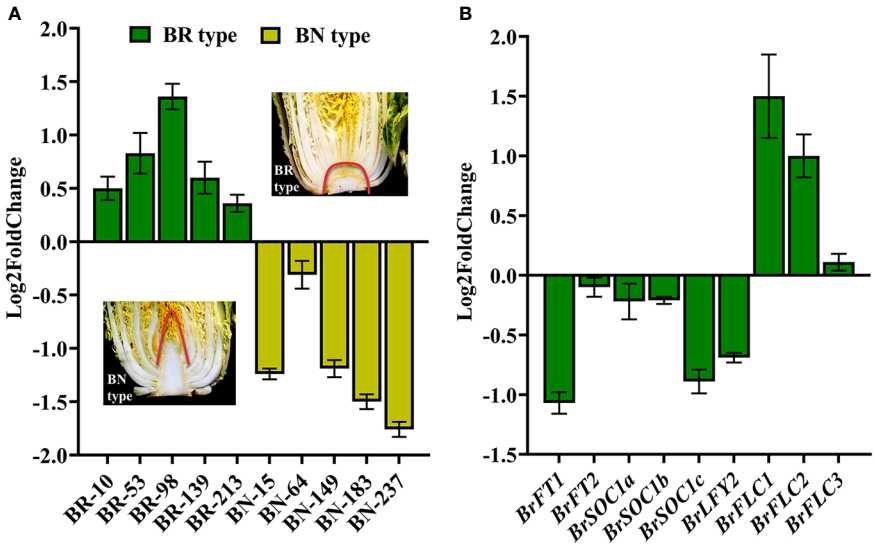


FIGURE 3
Expression patterns of *BR1* **(A)** and representative mark genes **(B)** in bolting resistant or non-resistant Chinese cabbage during heading stage. Examples of bolting resistant type (BR type) and bolting non-resistant type (BN type) Chinese cabbage were showed in the figure. Green column represented BR type, yellow column indicated BN type. BR type inbred lines 'BR-98' was used for expression analyses.

molecule activity, antioxidant activity) terms. KEGG pathway enrichment analysis also identified 47 significantly enriched KEGG pathways (Supplementary Table S6), with the top 20 for each DEG set being presented in Figure 5C. The majority of these DEGs were enriched in pathways including "metabolic pathways and biosynthesis of secondary metabolites (ko01100)", "plant-pathogen interaction (ko04626)", "phenylpropanoid biosynthesis (ko00940)", "plant MAPK signaling pathway (ko04016)", and "porphyrin and chlorophyll metabolism (ko00860)".

Next, DEGs associated with the floral transition pathway were compared between WT and *BR1OE* plants, revealing three MADS-box TFs among these DEGs including *AGAMOUS-LIKE 9* (*AtAGL9*, *AT1G24260*), *AtAP1*, and *AGAMOUS-LIKE 42* (*AtAGL42*, *AT5G62165*) (Figure 5D). These TFs were significantly downregulated in *BR1OE* plants relative to WT controls, in line with the delayed flowering time phenotype (Figure 4A). While not significantly downregulated, other floral integrator genes exhibited downward expression trends in *BR1OE* plants, including *AtSOC1*, *AtLFY*, and *FRUITFULL* (*AtFUL*, *AT5G60910*), whereas *AtFT* was expressed at low levels in both WT and mutant plants. The expression level of floral repressor *AtFLC* gene showed upward trends in *BR1OE* plants as expected. Interestingly, the downward expression trends of six key genes that involved in photoperiodic flowering pathway were observed in *BR1OE* plants, including *PHYTOCHROME A* (*AtPHYA*, *AT1G09570*), *PHYTOCHROME B* (*AtPHYB*, *AT2G18790*), *CRYPTOCHROME 1* (*AtCRY1*, *AT4G08920*), *CRYPTOCHROME 2* (*AtCRY2*, *AT1G04400*), *VASCULAR PLANT ONE ZINC FINGER PROTEIN 1* (*AtVOZ1*, *AT1G28520*), and *PHYTOCHROME INTERACTING FACTOR 4* (*AtPIF4*, *AT2G43010*). Moreover, the key chlorophyll biosynthesis-related gene *Protochlorophyllide oxidoreductase A* (*AtPORA*, *AT5G54190*) was also significantly downregulated in *BR1*OE mutant plants relative to WT controls, explaining the observed reduction in chlorophyll content (Figures 4G-I). These results indicated that *BR1* gene may regulate flowering primarily through vernalization and photoperiodic pathway, which further supported a model in which *BR1* serves as an important regulator of *Arabidopsis* floral transition.

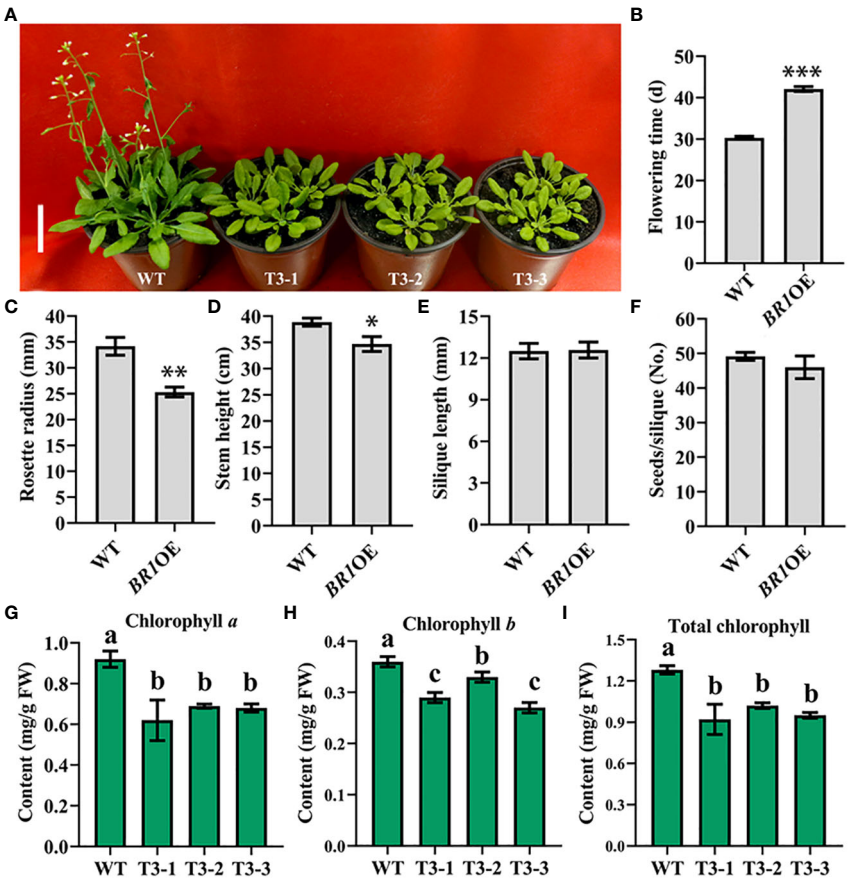


FIGURE 4
*BR1*OE mutant phenotypic characterization. **(A)** Under LD conditions the *BR1*OE transgenic plant exhibited a late-flowering phenotype. Representative images of 32-day-old transgenic plants from three separate genetic transformation events (T3-1, T3-2, and T3-3) and wild-type Col-0 (WT) are shown. Scale bar: 4 cm. **(B)** Flowering time. **(C)** Rosette radius. **(D)** Stem height. **(E)** Silique length. **(F)** Seed number per silique. Data are means ± SE of three independent measurements. Significantly differences were identified when comparing WT and *BR1*OE plants using Student's *t*-tests (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). **(G)** Chlorophyll *a* content. **(H)** Chlorophyll *b* content. **(I)** Total chlorophyll content. Analyses were performed with transgenic plants from three separate genetic transformation events (T3-1, T3-2, and T3-3) and wild-type Col-0 (WT) *Arabidopsis*. FW: fresh weight. Data are means ± SE of three independent measurements. Statistically distinct groups are marked with black letters (one-way ANOVA, $p < 0.05$).

3.5 BR1 localizes to the nucleus

A vector encoding the 35S::BR1::GFP fusion protein under the control of the CaMV35S promoter and a control empty 35S::GFP vector were prepared and used to conduct subcellular localization analyses. These assays revealed that BR1-GFP primarily localized to the nucleus, whereas control GFP signal was visible throughout the nuclear and cytosolic compartments (Figure 6). BR1-GRP and nuclear marker co-expression revealed complete overlap between these two signals, confirming that BR1 is a protein that localizes to the nucleus.

4 Discussion

4.1 *BR1* promotes improved bolting resistance

Genotypic analyses of these *OGs* across 524 accessions revealed four SNPs in the *BR1* gene, consistent with the high levels of *BR1* conservation in Chinese cabbage relative to other accessions. This provided preliminary support for the role of this and other *OGs* as regulators of flowering habits, offering insight into the distinct bolting mechanisms that may have evolved in *B. rapa* that are absent in *Arabidopsis* (Jiang et al., 2022). Potential *OG* and lineage-specific gene family contributions to species-specific phenotypes have previously been characterized in the context of multicellular development in dictyostelids (Luna and Chain, 2021). *OGs* are thus associated with species-specific traits, in line with the present results. Despite the close relationships between Chinese cabbage and *A. thaliana*, these plants have evolved certain distinct characteristics over the course of evolution including distinct leaf shape, leaf size, and flowering timing. During the heading stage of growth, Chinese cabbage plants exhibit a leafy head consisting of leaves with extremely inwardly curved blades on their shoot apex, with these leafy heads being composed of many heading leaves that generally curve following the rosette stage (Ren et al., 2018). *BR1* is mainly expressed in leaves, especially in the middle and inner leaves during the vegetative stage, suggesting that *BR1* may play a role in Chinese cabbage leafy head formation. *OGs* also exhibit a high degree of tissue-specific expression in different tissues and stages of development (Donoghue et al., 2011; Jiang et al., 2018), which may reflect the putative functions of *BR1* in the developmental regulation of Chinese cabbage. Well correspondence with delayed flowering phenotype in '*BR1*OE', the expression of *BR1* was

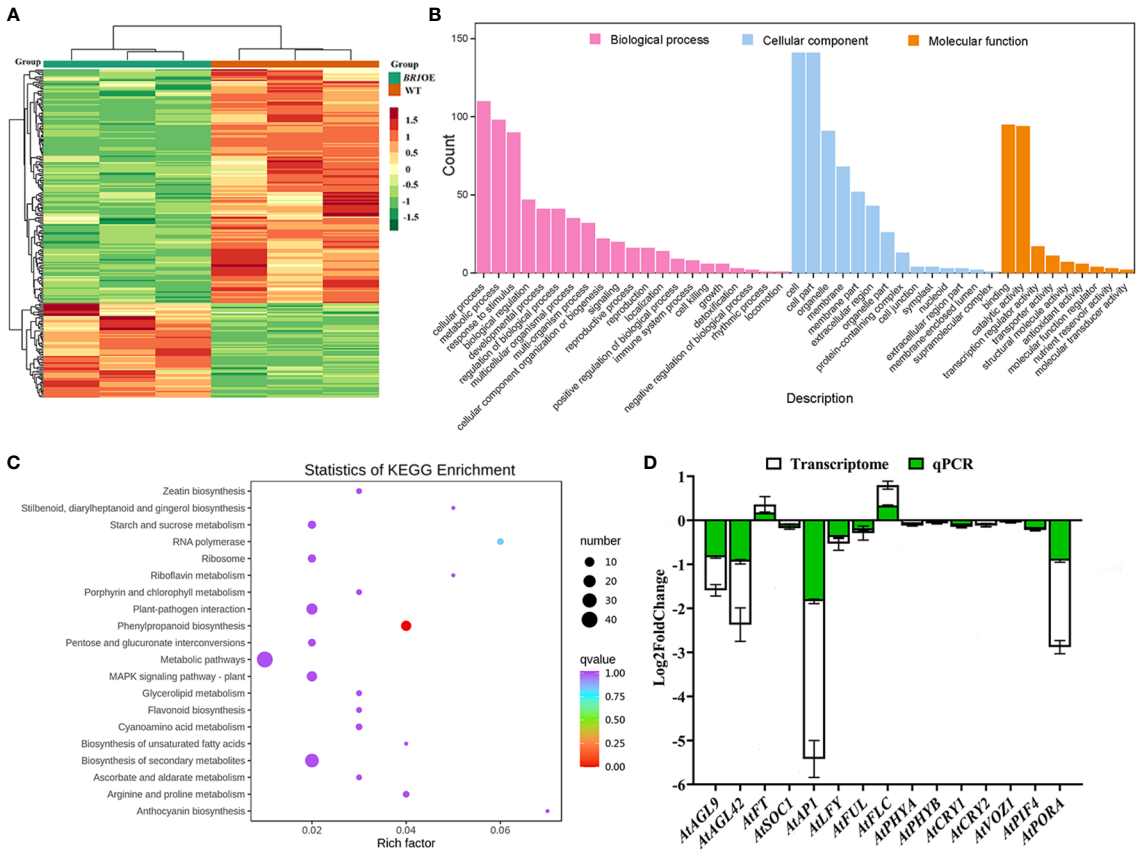


FIGURE 5
Identification of genes differentially expressed between WT and *BR1*OE plants. **(A)** DEG heatmap analysis. **(B)** DEG Gene Ontology (GO) classification was performed, with individual bars denoting the numbers of DEGs mapped to particular GO categories (pink: biological process; blue: cellular component; orange: molecular function). **(C)** The top 20 most enriched KEGG pathways associated with identified DEGs. Pathways are provided with the corresponding enrichment factors. **(D)** qPCR validation of gene expression profiles.

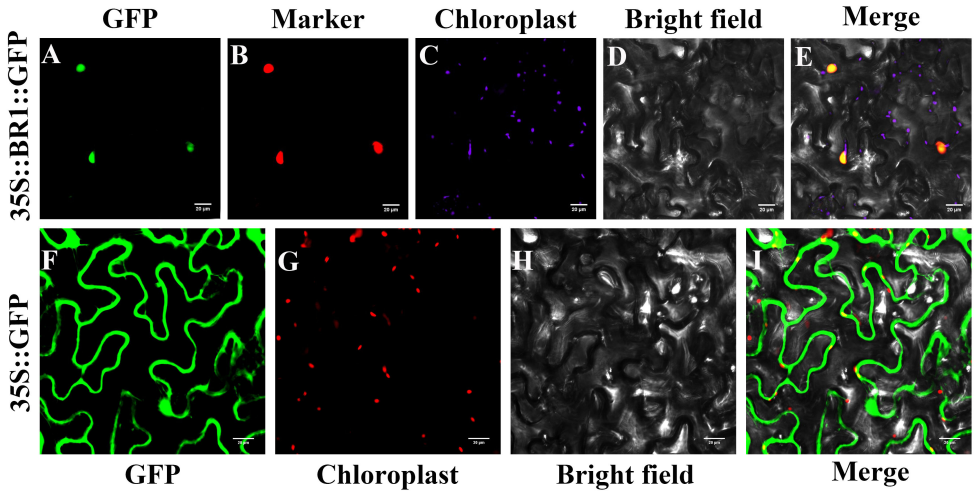


FIGURE 6
BR1 subcellular localization analyses. **(A, F)** GFP fluorescence. **(B)** Nuclear marker fluorescence. **(C, G)** chloroplast signals. **(D, H)** Brightfield images. **(E, I)** Merged images. A Leica confocal microscope was used to collect images at 48 h following agro-infiltration. Control GFP localization was evident throughout these cells. Scale bar: 20 μm.

increased in BR type, whereas decreased in BN type, which indicating that *BR1* may play a role in bolting resistance of Chinese cabbage. The expression levels of several mark genes were detected in representative bolting resistance inbred line, floral inhibition factor *BrFLCs* displayed upregulated expression, and other flowering integrators showed downregulated expression. Studies indicated that *FLC* homologues in *Brassica* species act similarly to *AtFLC* that can delay flowering (Yuan et al., 2009; Xi et al., 2018). The expression of integration factors was similar to previous study (Dong et al., 2016). These results further verified the bolting resistance phenotype of 'BR-98', and reflected the potential role of *BR1* in the regulation of bolting resistance.

4.2 Heterogeneous *BR1* expression regulates *Arabidopsis* bolting resistance

Recent years, bolting resistance of Chinese cabbage had attracted more and more attention from scientists. In this study, the *OG BR1* was identified as a novel regulator of plant morphological characteristics. *Arabidopsis BR1*OE mutants exhibited notably delayed flowering time together with significant decreases in stem height, rosette radius, and chlorophyll content. The overexpression of *GRAINS NUMBER 2* (*GN2*) gene unique to the 'Yuanjiang' common wild rice genome, was similarly found to promote a later heading date, decreased plant height, and a reduction in grain number relative to WT controls (Chen et al., 2017). *Brassica-Specific Genes 1* (*BSGs1*) has similarly been shown to be specifically expressed in Chinese cabbage during the heading stage, suggesting that it may be related to leafy head formation (Jiang et al., 2018). Flowering time is an agronomic trait that is critically important to Chinese cabbage production, and premature bolting can significantly reduce harvest yields and quality (Yuan et al., 2009). Chinese cabbage bolting and flowering generally necessitate vernalization and photoperiodism (Elers and Wiebe, 1984), and *BR1*-related flowering time delays may be linked to enhanced bolting resistance, highlighting a promising new pathway that can help prevent premature flowering for these economically important Chinese cabbage plants.

4.3 *BR1* alters flowering integrator gene expression in *Arabidopsis*

An appropriately timed floral transition is vital to ensuring (Srikanth and Schmid, 2011). By analyzing the transcriptome of the *Arabidopsis BR1*OE overexpression plants, we identified several flowering-related genes responsive to *BR1* overexpression. The transcriptomic sequencing and qPCR analyses performed herein confirmed that *BR1* overexpression in *Arabidopsis* significantly altered the expression of *AtAGL9*, *AtAGL42*, and *AtAP1*, in addition to promoting the downregulation of the floral transition identity genes *AtSOC1*, *AtLFY*, and *AtFUL*. Decreases in the expression of these integrator genes were consistent with observed flowering delays in mutant *BR1*OE plants. AtSOC1 can interact with FRUITFULL (AtFUL) to promote the activation of *AtLFY*, triggering the expression of the floral meristem identity gene *AtAP1* promoting floral commitment such that flowers develop on the SAM (Jaudal et al., 2018). *SPL* TFs can promote flowering time and floral fate, with MADS-box TFs including *AtAP1*, *AtFUL*, and *AtAGL9* serving as regulators of floral fate (Freytes et al., 2021). The *SOC1*-like gene, *AtAGL42*, is linked to the control of SAM and axillary meristem (AM) floral transition (Dorca-Fornell et al., 2011). AtAGL9 was shown to bind to and repress *AtSOC1*, while activating expression of a large number of floral homeotic genes

(Srikanth and Schmid, 2011). The expression level of *AtFLC* gene displayed upward trends in *BR1OE* plants as expected. As a major inhibitor of flowering, FLC suppressed the expression of transcription factors (TFs) needed for the SAM floral transition (Cho et al., 2017). Low expression level in *AtFT* was also evident in *BR1*OE plants relative to WT controls, and this phenomenon still needs further clarification. Eight core regulatory genes in photoperiodic flowering pathway showed downward expression trends, including *AtPHYA*, *AtPHYB*, *AtCRY1*, *AtCRY2*, *AtVOZ1*, and *AtPIF4*. These genes regulated flowering, and both of these mutants flower late under different conditions (Simpson and Dean, 2002; Thines et al., 2014). How the overexpression of *BR1* gene affects the expression of these genes still needs further study.

Notably, *BR1* overexpression resulted in significant *AtPORA* downregulation in *Arabidopsis*, consistent with the observed reductions in chlorophyll levels in *BR1*OE transgenic plants (Figure 4G, H, I). POR is a key enzyme for the light-induced greening of etiolated angiosperm plants, which is one of three known light-dependent enzymes, catalyses reduction of the photosensitizer and substrate protochlorophyllide to form the pigment chlorophyllide (Reinbothe et al., 2010; Buhr et al., 2017; Zhang et al., 2019). The *Arabidopsis porA-1* mutant plants have been shown to exhibit severe photoautotrophic growth defects and decreases in total chlorophyll content (Paddock et al., 2012), confirming the present results. Through comparisons of transcriptomic profiles, most enriched KEGG pathways were metabolic pathways related to the biosynthesis of secondary metabolites. GO enrichment analyses indicated that all DEGs were related to metabolic, developmental, reproductive, and growth-related processes. Based on the above analysis, we speculated that *BR1* may regulate flowering time delay through vernalization and photoperiodic pathway. However, the mechanisms through which the *BR1* pathway contributes to delayed flowering warrant further research. As BR1 localizes to the nuclear compartment, the identification of upstream regulatory TFs and downstream target proteins can be readily performed through respective yeast one-hybrid and yeast two-hybrid screens. Further knockout and overexpression analyses in Chinese cabbage will thus be essential to fully characterize how *BR1* regulates bolting resistance.

5 Conclusions

The highly conserved nature of *BR1* in Chinese cabbage accessions supports its potential role as a regulator of bolting resistance. The expression of *BR1* was increased in bolting resistant Chinese cabbage and decreased in bolting non-resistant type. The overexpression of *BR1* in *Arabidopsis* resulted in delayed flowering time and other obvious phenotypes. Both transcriptomic sequencing and qPCR analyses further revealed that nuclear-located orphan gene *BR1* may function as a new mediator of flowering time through vernalization and photoperiodic pathway. These results offer novel insight into the links between *OGs* and flowering time in *B. rapa*, providing a theoretical basis for future studies aimed at examining the mechanisms governing bolting resistance in this economically important species.

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: PRJNA922732 (SRA).

Author contributions

MJ and HL conceived and designed the experiments. MJ, YZ, and HL performed the experiments. MJ, YZ, XY, XL, and HL analyzed the data. MJ and HL drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the Doctoral Initiating Fund Project of Jilin Agricultural Science and Technology College (Grant No. 20215019).

Acknowledgments

We thank Prof. Zhongyun Piao, Dr. Wenxing Pang and Dr. Zongxiang Zhan (College of Horticulture, Shenyang Agricultural University) for their kindly help on this study.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1135684/full#supplementary-material

References

Bao, S., Hua, C., Shen, L., and Yu, H. (2020). New insights into gibberellin signaling in regulating flowering in *Arabidopsis*. *J. Integr. Plant Biol.* 62, 118–131. doi: 10.1111/jipb.12892

Blümel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops-what did we learn from *Arabidopsis*? *Curr. Opin. Biotechnol.* 32, 121–129. doi: 10.1016/j.copbio.2014.11.023

Brennan, C. J., Zhou, B., Benbow, H. R., Ajaz, S., Karki, S. J., Hehir, J. G., et al. (2020). Taxonomically restricted wheat genes interact with small secreted fungal proteins and enhance resistance to septoria tritici blotch disease. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00433

Buhr, F., Lahroussi, A., Springer, A., Rustgi, S., von Wettstein, D., Reinbothe, C., et al. (2017). NADPH:protochlorophyllide oxidoreductase b (PORB) action in *Arabidopsis thaliana* revisited through transgenic expression of engineered barley PORB mutant proteins. *Plant Mol. Biol.* 94, 45–59. doi: 10.1007/s11103-017-0592-x

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, 166. doi: 10.1186/s13059-021-02383-2

Cai, J. J., Woo, P. C., Lau, S. K., Smith, D. K., and Yuen, K. Y. (2006). Accelerated evolutionary rate may be responsible for the emergence of lineage-specific genes in ascomycota. *J. Mol. Evol.* 63, 1–11. doi: 10.1007/s00239-004-0372-5

Chen, H., Tang, Y., Liu, J., Tan, L., Jiang, J., Wang, M., et al. (2017). Emergence of a novel chimeric gene underlying grain number in rice. *Genetics* 205, 993–1002. doi: 10.1534/genetics.116.188201

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat. Genet.* 48, 1218–1224. doi: 10.1038/ng.3634

Cheng, J. Z., Zhou, Y. P., Lv, T. X., Xie, C. P., and Tian, C. E. (2017). Research progress on the autonomous flowering time pathway in *Arabidopsis*. *Physiol. Mol. Biol. Plants* 23, 477–485. doi: 10.1007/s12298-017-0458-3

Cho, L. H., Yoon, J., and An, G. (2017). The control of flowering time by environmental factors. *Plant J.* 90, 708–719. doi: 10.1111/tpj.13461

Creux, N., and Harmer, S. (2019). Circadian rhythms in plants. *Cold Spring Harb. Perspect. Biol.* 11, a034611. doi: 10.1101/cshperspect.a034611

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D. (2015). Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Mol. Plant* 8, 935–945. doi: 10.1016/j.molp.2014.12.008

Dong, X., Yi, H., Han, C. T., Nou, I. S., Swaraz, A. M., and Hur, Y. (2016). Genome-wide analysis of genes associated with bolting in heading type Chinese cabbage. *Euphytica* 212, 65–82. doi: 10.1007/s10681-016-1759-2

Donoghue, M. T., Keshavaiah, C., Swamidatta, S. H., and Spillane, C. (2011). Evolutionary origins of brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol. Biol.* 11, 47. doi: 10.1186/1471-2148-11-47

Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., and Kater, M. M. (2011). The *Arabidopsis SOC1*-like genes *AGL42*, *AGL71* and *AGL72* promote flowering in the shoot apical and axillary meristems. *Plant J.* 67, 1006–1017. doi: 10.1111/j.1365-313X.2011.04653.x

Dossa, K., Zhou, R., Li, D., Liu, A., Qin, L., Mmadi, M. A., et al. (2021). A novel motif in the 5'-UTR of an orphan gene '*Big root biomass*' modulates root biomass in sesame. *Plant Biotechnol. J.* 19, 1065–1079. doi: 10.1111/pbi.13531

Elers, B., and Wiebe, H. J. (1984). Flower formation of Chinese cabbage. i. response to vernalization and photoperiods. *Sci. Hortic.* 22, 219–231. doi: 10.1016/0304-4238(84)90055-4

Fang, H., Shen, S., Wang, D., Zhang, F., Zhang, C., Wang, Z., et al. (2021). A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. *Sci. Bull.* 66, 2381–2393. doi: 10.1016/j.scib.2021.06.014

Freytes, S. N., Canelo, M., and Cerdán, P. D. (2021). Regulation of flowering time: When and where? *Curr. Opin. Plant Biol.* 63, 102049. doi: 10.1016/j.pbi.2021.102049

Fu, W., Huang, S., Gao, Y., Zhang, M., Qu, G., Wang, N., et al. (2020). Role of *BrSDG8* on bolting in Chinese cabbage (*Brassica rapa*). *Theor. Appl. Genet.* 133, 2937–2948. doi: 10.1007/s00122-020-03647-4

Izawa, T. (2021). What is going on with the hormonal control of flowering in plants? *Plant J.* 105, 431–445. doi: 10.1111/tpj.15036

Jaudal, M., Zhang, L., Che, C., Li, G., Tang, Y., Wen, J., et al. (2018). A *SOC1-like* gene *MtSOC1a* promotes flowering and primary stem elongation in medicago. *J. Exp. Bot.* 69, 4867–4880. doi: 10.1093/jxb/ery284

Jiang, M., Dong, X., Lang, H., Pang, W., Zhan, Z., Li, X., et al. (2018). Mining of *Brassica*-specific genes (*BSGs*) and their induction in different developmental stages and under *Plasmodiophora brassicae* stress in *Brassica rapa*. *Int. J. Mol. Sci.* 19, 2064. doi: 10.3390/ijms19072064

Jiang, C., Hei, R., Yang, Y., Zhang, S., Wang, Q., Wang, W., et al. (2020a). An orphan protein of *Fusarium graminearum* modulates host immunity by mediating proteasomal degradation of TaSnRK1α. *Nat. Commun.* 11, 4382. doi: 10.1038/s41467-020-18240-y

Jiang, M., Li, X., Dong, X., Zu, Y., Zhan, Z., Piao, Z., et al. (2022). Research advances and prospects of orphan genes in plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.947129

Jiang, M., Zhan, Z., Li, H., Dong, X., Cheng, F., and Piao, Z. (2020b). *Brassica rapa* orphan genes largely affect soluble sugar metabolism. *Hortic. Res.* 7, 181. doi: 10.1038/s41438-020-00403-z

Johansson, M., and Staiger, D. (2015). Time to flower: Interplay between photoperiod and the circadian clock. *J. Exp. Bot.* 66, 719–730. doi: 10.1093/jxb/eru441

Jung, C., and Müller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14, 563–573. doi: 10.1016/j.tplants.2009.07.005

Kinoshita, A., and Richter, R. (2020). Genetic and molecular basis of floral induction in *Arabidopsis thaliana*. *J. Exp. Bot.* 71, 2490–2504. doi: 10.1093/jxb/eraa057

Li, L., Foster, C. M., Gan, Q., Nettleton, D., James, M. G., Myers, A. M., et al. (2009). Identification of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis* leaves. *Plant J.* 58, 485–498. doi: 10.1111/j.1365-313X.2009.03793.x

Li, L., Li, X., Liu, Y., and Liu, H. (2015a). Flowering responses to light and temperature. *Sci. China Life Sci.* 59, 403–408. doi: 10.1007/s11427-015-4910-8

Li, Q., Peng, A., Yang, J., Zheng, S., Li, Z., Mu, Y., et al. (2022). A 215-bp indel at intron I of *BoFLC2* affects flowering time in *Brassica oleracea* var. *capitata* during vernalization. *Theor. Appl. Genet.* 135, 2785–2797. doi: 10.1007/s00122-022-04149-1

Li, G., Wu, X., Hu, Y., Muñoz-Amatriaín, M., Luo, J., Zhou, W., et al. (2019). Orphan genes are involved in drought adaptations and ecoclimatic-oriented selections in domesticated cowpea. *J. Exp. Bot.* 70, 3101–3110. doi: 10.1093/jxb/erz145

Li, L., and Wurtele, E. S. (2015). The *QQS* orphan gene of *Arabidopsis* modulates carbon and nitrogen allocation in soybean. *Plant Biotechnol. J.* 13, 177–187. doi: 10.1111/pbi.12238

Li, L., Zheng, W., Zhu, Y., Ye, H., Tang, B., Arendsee, Z. W., et al. (2015b). QQS orphan gene regulates carbon and nitrogen partitioning across species *via* NF-YC interactions. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14734–14739. doi: 10.1073/pnas.1514670112

Lin, S. I., Wang, J. G., Poon, S. Y., Su, C. L., Wang, S. S., and Chiou, T. J. (2005). Differential regulation of *FLOWERING LOCUS c* expression by vernalization in cabbage and *Arabidopsis*. *Plant Physiol.* 137, 1037–1048. doi: 10.1104/pp.104.058974

Luna, S. K., and Chain, F. J. J. (2021). Lineage-specific genes and family expansions in dictyostelid genomes display expression bias and evolutionary diversification during development. *Genes (Basel)* 12, 1628. doi: 10.3390/genes12101628

Ma, D., Ding, Q., Guo, Z., Zhao, Z., Wei, L., Li, Y., et al. (2021). Identification, characterization and expression analysis of lineage-specific genes within mangrove species *Aegiceras corniculatum*. *Mol. Genet. Genomics* 296, 1235–1247. doi: 10.1007/s00438-021-01810-0

Ma, D., Lai, Z., Ding, Q., Zhang, K., Chang, K., Li, S., et al. (2022). Identification, characterization and function of orphan genes among the current cucurbitaceae genomes. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.872137

Ma, S., Yuan, Y., Tao, Y., Jia, H., and Ma, Z. (2020). Identification, characterization and expression analysis of lineage-specific genes within *Triticeae*. *Genomics* 112, 1343–1350. doi: 10.1016/j.ygeno.2019.08.003

Mero, C. E., and Honma, S. (1984). Inheritance of bolt resistance in an interspecific cross of *Brassica* species: II. chikale (*B. campestris* l. ssp. *pekinensis* × *B. napus* l.) × Chinese cabbage. *J. Hered.* 75, 6485–6487. doi: 10.1093/oxfordjournals.jhered.a109991

Moon, H., Jeong, A. R., Kwon, O. K., and Park, C. J. (2022). Oryza-specific orphan protein triggers enhanced resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.859375

Ni, F., Qi, J., Hao, Q., Lyu, B., Luo, M. C., Wang, Y., et al. (2017). Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species. *Nat. Commun.* 8, 15121. doi: 10.1038/ncomms15121

O'Conner, S., Neudorf, A., Zheng, W., Qi, M., Zhao, X., Du, C., et al. (2018). "From arabidopsis to crops: the arabidopsis QQS orphan gene modulates nitrogen allocation across species," in *Engineering nitrogen utilization in crop plants*. Eds. A. Shrawat, A. Zayed and D. A. Lightfoot (Switzerland, AG: Springer Cham), 95–117. doi: 10.1007/978-3-319-92958-3_6

Paddock, T., Lima, D., Mason, M. E., Apel, K., and Armstrong, G. A. (2012). *Arabidopsis* light-dependent protochlorophyllide oxidoreductase a (PORA) is essential for normal plant growth and development. *Plant Mol. Biol.* 78, 447–460. doi: 10.1007/s11103-012-9873-6

Qi, M., Zheng, W., Zhao, X., Hohenstein, J. D., Kandel, Y., O'Conner, S., et al. (2019). QQS orphan gene and its interactor NF-YC4 reduce susceptibility to pathogens and pests. *Plant Biotechnol. J.* 17, 252–263. doi: 10.1111/pbi.12961

Razi, H., Howell, E. C., Newbury, H. J., and Kearsey, M. J. (2008). Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in

Brassica oleracea? *Theor. Appl. Genet.* 116, 179–192. doi: 10.1007/s00122-007-0657-3

Reinbothe, C., El Bakkouri, M., Buhr, F., Muraki, N., Nomata, J., Kurisu, G., et al. (2010). Chlorophyll biosynthesis: Spotlight on protochlorophyllide reduction. *Trends Plant Sci.* 15, 614–624. doi: 10.1016/j.tplants.2010.07.002

Ren, W., Wang, H., Bai, J., Wu, F., and He, Y. (2018). Association of microRNAs with types of leaf curvature in *Brassica rapa*. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00073

Shen, S., Peng, M., Fang, H., Wang, Z., Zhou, S., Jing, X., et al. (2021). An *Oryza*-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci. Bull.* 66, 2369–2380. doi: 10.1016/j.scib.2021.03.015

Simpson, G. G., and Dean, C. (2002). *Arabidopsis*, the Rosetta stone of flowering time? *Science* 296, 285–289. doi: 10.1126/science.296.5566.285

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y

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

Tanvir, R., Ping, W., Sun, J., Cain, M., Li, X., and Li, L. (2022a). *AtQQS* orphan gene and *NtNF-YC4* boost protein accumulation and pest resistance in tobacco (*Nicotiana tabacum*). *Plant Sci.* 317, 111198. doi: 10.1016/j.plantsci.2022.111198

Tanvir, R., Wang, L., Zhang, A., and Li, L. (2022b). Orphan genes in crop improvement: enhancing potato tuber protein without impacting yield. *Plants (Basel)* 11, 3076. doi: 10.3390/plants11223076

Thines, B. C., Youn, Y., Duarte, M. I., and Harmon, F. G. (2014). The time of day effects of warm temperature on flowering time involve PIF4 and PIF5. *J. Exp. Bot.* 65, 1141–1151. doi: 10.1093/jxb/ert487

Wang, J. W. (2014). Regulation of flowering time by the miR156-mediated age pathway. *J. Exp. Bot.* 65, 4723–4730. doi: 10.1093/jxb/eru246

Wang, C., Chen, S., Feng, A., Su, J., Wang, W., Feng, J., et al. (2021). *Xa7*, a small orphan gene harboring promoter trap for *AvrXa7*, leads to the durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Rice* 14, 48. doi: 10.1186/s12284-021-00490-z

Wang, Z., Li, J., Chen, S., Heng, Y., Chen, Z., Yang, J., et al. (2017). Poaceae-specific *MS1* encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl. Acad. Sci. U.S.A.* 114, 12614–12619. doi: 10.1073/pnas.1715570114

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

Xin, X., Su, T., Li, P., Wang, W., Zhao, X., Yu, Y., et al. (2020). A histone H4 gene prevents drought-induced bolting in Chinese cabbage by attenuating the expression of flowering genes. *J. Exp. Bot.* 72, 623–635. doi: 10.1093/jxb/eraa452

Yadeta, K. A., Valkenburg, D. J., Hanemian, M., Marco, Y., and Thomma, B. P. (2014). The brassicaceae-specific *EWR1* gene provides resistance to vascular wilt pathogens. *PloS One* 9, e88230. doi: 10.1371/journal.pone.0088230

Yuan, Y. X., Wu, J., 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

Zhan, Z., Liu, H., Yang, Y., Liu, S., Li, X., and Piao, Z. (2022). Identification and characterization of putative effectors from *Plasmodiophora brassicae* that suppress or induce cell death in *Nicotiana benthamiana*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.881992

Zhang, X., Feng, C., Wang, M., Li, T., Liu, X., and Jiang, J. (2021). Plasma membrane-localized SlSWEET7a and SlSWEET14 regulate sugar transport and storage in tomato fruits. *Hortic. Res.* 8, 186. doi: 10.1038/s41438-021-00624-w

Zhang, S., Heyes, D. J., Feng, L., Sun, W., Johannissen, L. O., Liu, H., et al. (2019). Structural basis for enzymatic photocatalysis in chlorophyll biosynthesis. *Nature* 574, 722–725. doi: 10.1038/s41586-019-1685-2

Zhang, X., Ma, W., Liu, M., Li, X., Li, J., Lu, Y., et al. (2022). OCTOPUS regulates BIN2 to control leaf curvature in Chinese cabbage. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2208978119. doi: 10.1073/pnas.2208978119

Zhao, D. S., Li, Q. F., Zhang, C. Q., Zhang, C., Yang, Q. Q., Pan, L. X., et al. (2018). *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat. Commun.* 9, 1240. doi: 10.1038/s41467-018-03616-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1153000

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Kai Zhang,
Southwest University, China
Zhansheng Li,
Institute of Vegetables and Flowers (CAAS), China
Xiaonan Li,
Shenyang Agricultural University, China

*CORRESPONDENCE
Jihong Hu
hujh05@nwafu.edu.cn
Shuai Liu
SLiu@cc.hawaii.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 28 January 2023
ACCEPTED 21 March 2023
PUBLISHED 14 April 2023

CITATION
Zhang C, Gong R, Zhong H, Dai C, Zhang R, Dong J, Li Y, Liu S and Hu J (2023) Integrated multi-locus genome-wide association studies and transcriptome analysis for seed yield and yield-related traits in *Brassica napus*. *Front. Plant Sci.* 14:1153000. doi: 10.3389/fpls.2023.1153000

Integrated multi-locus genome-wide association studies and transcriptome analysis for seed yield and yield-related traits in *Brassica napus*

Cuiping Zhang[1†], Ruolin Gong[1†], Hua Zhong[2†], Chunyan Dai[1], Ru Zhang[1], Jungang Dong[1], Yangsheng Li[3], Shuai Liu[2*] and Jihong Hu[1*]

[1]State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, China, [2]Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii at Manoa, Honolulu, HI, United States, [3]State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, Wuhan, China

Rapeseed (*Brassica napus* L.), the third largest oil crop, is an important source of vegetable oil and biofuel for the world. Although the breeding and yield has been improved, rapeseed still has the lowest yield compared with other major crops. Thus, increasing rapeseed yield is essential for the high demand of vegetable oil and high-quality protein for live stocks. Silique number per plant (SN), seed per pod (SP), and 1000-seed weight (SW) are the three important factors for seed yield in rapeseed. Some yield-related traits, including plant height (PH), flowering time (FT), primary branch number (BN) and silique number per inflorescence (SI) also affect the yield per plant (YP). Using six multi-locus genome-wide association study (ML-GWAS) approaches, a total of 908 yield-related quantitative trait nucleotides (QTNs) were identified in a panel consisting of 403 rapeseed core accessions based on whole-genome sequencing. Integration of ML-GWAS with transcriptome analysis, 79 candidate genes, including *BnaA09g39790D* (*RNA helicase*), *BnaA09g39950D* (*Lipase*) and *BnaC09g25980D* (*SWEET7*), were further identified and twelve genes were validated by qRT-PCRs to affect the SW or SP in rapeseed. The distribution of superior alleles from nineteen stable QTNs in 20 elite rapeseed accessions suggested that the high-yielding accessions contained more superior alleles. These results would contribute to a further understanding of the genetic basis of yield-related traits and could be used for crop improvement in *B. napus*.

KEYWORDS

rapeseed, yield, seed weight, multi-locus GWAS, candidate gene, RNA-seq

Introduction

Brassica napus (*B. napus*, AACC, 2n = 38) is one of the most important oilseed crops worldwide as vegetable oil, animal feed and biofuel (Lu et al., 2019; Song et al., 2020). However, the deteriorating environment and lack of arable land make the yield of rapeseed insufficient to support the demand. Therefore, increasing rapeseed yields is a research priority for rapeseed breeders to meet the future demand of oilseed rape production. The yield of rapeseed is a complex quantitative trait and mainly determined by three yield-component traits, including 1000-seed weight (SW), silique number per plant (SN) and seed per pod (SP) (Zhao et al., 2016; Lu et al., 2017). Seed yield is also influenced by yield-related traits such as plant height (PH), flowering time (FT), primary branch number (BN), length of main inflorescence, and silique number of main inflorescence (SI) in rapeseed (Shi et al., 2009; Zhao et al., 2016). There is a correlation among these yield traits, and they interact with each other to jointly determine the rapeseed yield. In addition, the relationships between these traits are intricate, for example, SW was positively correlated with yield per plant, plant height and length of main inflorescence, while it was negatively correlated with seed number per pod and flowering time. Seed size affects the SW and SP, which is of great value for crop improvement in rapeseed (Li et al., 2019a). Thus, dissecting the genetic basis and molecular mechanism of yield traits will facilitate and accelerate breeding programs for yield in rapeseed.

Rapeseed yield component and related traits are all complex quantitative traits governed by multiple genes. In previous studies, some loci or genes for yield traits in rapeseed were identified using quantitative trait locus (QTL) mapping and map-based cloning (Shi et al., 2009; Li et al., 2015; Liu et al., 2015; Li et al., 2019b; Shi et al., 2019; Wang et al., 2021a). Zhou et al. (2014) mapped 736 QTLs associated with the yield in the A and C subgenomes of *B. napus*, which were distributed over 19 chromosomes, mostly located on A03. Raboanatahiry et al. (2018) detected 972 QTLs associated with seed yield and yield-related traits in *B. napus*, identifying 147 potential candidate genes that could affect nine different traits. With the development of high-density customized single nucleotide polymorphism (SNPs), genome wide association study (GWAS) has become a powerful tool for deciphering the genetic architecture of complex quantitative traits (Lu et al., 2017; Zhong et al., 2021a). Based on 33,186 genomic SNPs from the 60 K *Brassica* Illumina Infinium SNP array, a new QTL was fine mapped onto chromosome C03 in *B. napus* and a gene controlling the branching number phenotype was identified (He et al., 2017). Integrated GWAS and transcriptome analysis, auxin-related genes were identified to associate with leaf petiole angle at the seedling stage in *B. napus* (Hu et al., 2021a). And stable QTLs localized on chromosomes A07, A09, and C08 were identified for silique length (SL) using GWAS combined with RNA-seq (Wang et al., 2021a). With the advance of next-generation sequencing (NGS) technology, mega-level SNPs or genetic variations *via* whole genome resequencing in population could detect more loci for the yield traits in crops (Zhang et al., 2022a). Using 10,658 high-quality SNP markers, a total of 497 SNPs were detected to associate with yield-related traits in *B. napus* (Zhang et al., 2022b). Based on high-quality 670,028 SNPs, GWASs were conducted using multi-locus random mixed linear model for 11 important traits in rapeseed (Lu et al., 2019). Using a nested association mapping population, SNP-GWAS, and presence and absence variation (PAV)-GWAS identified loci and structural variations for silique length, SW and FT (Song et al., 2020). Furthermore, 56 agronomically traits, including plant architecture and yield traits were examined using whole-genome resequencing in 403 diverse rapeseed accessions by GWAS and identified 26 loci associated with SW and silique length (Hu et al., 2022). Most of these GWAS studies used the single-locus GWAS (SL-GWAS) methods, such as the general linear model (GLM), the mixed linear model (MLM), efficient mixed-model association eXpedited (EMMAX), and factored spectrally transformed linear mixed models (FaST-LMM) (Zhong et al., 2021b). However, the SL-GWAS methods rule out many significant loci, including minor effect loci due to their highly stringent Bonferroni correction (Wang et al., 2016).

Recently, multi-locus GWAS (ML-GWAS) methodologies, including FASTmrEMMA, ISIS EM-BLASSO, mrMLM, FASTmrMLM, pLARmEB, pKWmEB, were developed as an effective approach for association analysis (Cui et al., 2018; Zhang et al., 2020). And various combinations of ML-GWAS methods have proven to be significantly effective in controlling false positive rates (Misra et al., 2017; Zhong et al., 2021c). Thus, ML-GWAS has been used to discover novel quantitative trait nucleotides (QTNs) in many crops. In wheat, five ML-GWAS models were utilized to successfully identify new QTL for yield-related traits in 272 local Chinese wheat landraces based on 172,711 SNPs (Lin et al., 2021). ML-GWAS of 144 maize inbred lines genotyped with 43,427 SNPs identified a large number of significant QTNs and 40 candidate genes associated with the regenerative capacity of the embryonic callus (Ma et al., 2018). In soybean, 129 significant QTNs related to protein content were identified by five ML-GWAS methods, and 8 candidate genes were predicted to be involved in protein synthesis and metabolism (Zhang et al., 2018). Using ML-GWAS, 74 significant QTN hotspots have been identified to associate with five yield-related traits in rice (Zhong et al., 2021c). However, ML-GWAS methods for yield-related traits in rapeseed have not yet been performed, especially based on the whole genome resequencing data. Given the efficiency and various models of ML-GWAS, more novel loci and candidate genes could be identified to associate with yield traits.

In the present study, we aimed to identify novel loci for seed yield and yield-related traits in rapeseed. Six ML-GWAS approaches were used to determine novel QTNs based on high-quality SNPs in 403 rapeseed accessions for eight yield traits in three environments. And we also analyzed the significant QTNs and pinpointed multiple candidate causal genes for the QTNs. The candidate genes and elite alleles identified with yield traits will provide an insight into further exploration of the genetic architecture of yield traits in rapeseed and genetic improvement of rapeseed.

Materials and methods

Plant materials and phenotype evaluation

A highly diverse natural population consisting of 403 core rapeseed germplasms were used as previous study (Hu et al., 2022). The set of rapeseed accessions includes spring (102), semi-winter (179) and winter (129) types from China, Germany, France, Canada, Japan, USA and other countries. And phenotyping of the 403 accessions were evaluated under three environments: E1, Yangluo (30.38° N, 114.50° E) in 2013, E2, Nanchang (28.37° N, 116.27° E) in 2014, and E3, Wuhan (30.58° N, 113.68° E) in 2016, which were download from https://www.cgris.net/rapedata/ or http://brassicanapusdata.cn/. The yield traits including SI, SN, SP, SW, and YP as well as three yield-related traits, such as PH, BN, and FT were measured according to the measurement standards (Chen et al., 2014; Li et al., 2016). The large seed size accession R01 and small seed size accession R56 were grown in a greenhouse (light/dark 16/8 h photoperiod and 25°C/20°C day/night temperature) in Northwest Agriculture and Forestry University, Yangling, Shaanxi, China.

Genotyping data processing

Whole genome resequencing was performed on the Illumina HiSeq 4000 platform with 150 bp paired-end and download from NCBI (PRJNA416679) (Hu et al., 2022). GATK (v.3.3) was used for SNP calling and then excluded SNP calling errors, retaining only high-quality SNPs (minor allele frequency ≥ 0.05, miss ≤ 0.2 and sequencing depth ≥ 6) for subsequent analysis (McKenna et al., 2010). These SNPs were processed by PLINK 1.9 with parameter –maf 0.05 –geno 0.05, –snps-only and imputed by Beagele 5.0 (Browning et al., 2018; Zhong et al., 2021c). Finally, a total of 7,531,945 high-quality SNPs were employed for GWAS analysis.

Multi locus-GWAS analysis

Six ML-GWAS methods, including the mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO, were used to identify significant QTN. All six ML-GWAS approaches are implemented in R package "mrMLM" (https://cran.r-project.org/web/packages/mrMLM/index.html) (Wang et al., 2016). Default values were used for all parameters and a threshold of logarithm of odds (LOD ≥ 3 or $P \leq 0.0002$) was chosen to examine the association between markers and yield-related traits (Zhang et al., 2019). Principal component analysis and kinship matrices were used in all methods. The R package CMplot (https://github.com/yinliLin/R-CMplot) was employed to visualize Manhattan and QQ plots of GWAS. Using the Tassel 5.2 tool (Bradbury et al., 2007), LDs between SNPs were estimated as the squared correlation coefficient (R^2) of alleles, and R^2 values were calculated within a 0 to 10 cM window. The phenotypic-effect value of allelic variation for each trait was calculated by the phenotypic data across the 403 accessions. The relative phenotypic data were visualized with box plots using the R 4.2.1 software.

Identification of candidate genes

The putative candidate loci were identified by at least two different GWAS methods. Candidate genes were predicted from the upstream or downstream 200 kb region of stable QTL loci using the 'Darmor-*bzh*' reference genome (https://www.genoscope.cns.fr/brassicanapus) (Chalhoub et al., 2014). Then the candidate genes were further annotated on NCBI (https://www.ncbi.nlm.nih.gov/) and annotated for *Arabidopsis* homologous genes by BLAST analysis (Hu et al., 2021a). Haplotype analysis of the QTN association regions across the 403 rapeseed accessions was conducted by Haploview v4.2 (Barrett et al., 2005).

Transcriptome analysis

Total RNA was extracted from developing siliques (include seeds) of two extremely SW accessions R01 (large seed) and R56 (small seed) for RNA-seq at two weeks after pollination (2 WAP), three weeks after pollination (3 WAP), and four weeks after pollination (4 WAP). Illumina's NEBNext® UltraTM RNA Library Preparation Kit was used to construct libraries and quality control was checked by Agilent 2100 Bioanalyzer System. Clean reads after filtering the raw reads were mapped to the "Darmor-*bzh*" reference genome (https://www.genoscope.cns.fr/brassicanapus) using HISAT2 software (Chalhoub et al., 2014; Kim et al., 2015). Gene expression levels were normalized using the FPKM (fragments per kilobase per million reads) values by StringTie (Pertea et al., 2015). Differential expression analysis between the sample pairs in two rapeseed accessions was conducted by DESeq2 (Love et al., 2014). Differentially expressed genes (DEGs) were determined with false discovery rate (FDR) < 0.05 and |log2 (fold change)| ≥ 1. GO (Gene ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of the DEGs were performed using AgiGO 2.0 and KOBAS3.0, respectively (Tian et al., 2017; Bu et al., 2021).

Candidate gene expression analysis

Candidate genes were predicted based on the ML-GWAS and transcriptome analysis, and were further validated by quantitative real-time PCR (qRT-PCR). Briefly, 1μg of total RNA was used for RNA-seq and cDNA synthesis was performed using HiScript®Q RT SuperMix (Vazyme, China). Data collection was performed in QuantStudio™ real-time PCR software (Thermo Fisher Scientific, Waltham, MA, USA). All the primers are listed in Supplementary Table S1. Data were normalized by the internal control gene *BnACTIN* (*BnaA03g55890D*) and relative expression levels were calculated using a $2^{-\Delta\Delta CT}$ analysis method (Livak and Schmittgen, 2001).

Results

Phenotypic evaluation for yield traits

Phenotypic values for eight yield traits of rapeseed in three environments, including PH, FT, BN, SI, SN, SP, SW, and YP were used to determine whether significant phenotypic differences existed in these traits (Supplementary Figure S1). The phenotypic assessments revealed a wide range of variation between the different accessions, with the frequency distribution of all traits approximating a normal distribution (Figure 1). In addition, we noted that yield-related traits were differentially affected by environments, with SW and SP remaining relatively stable across environments, while PH and BN were more variable. Taken together, the extent of available variation for the different traits suggested that the set of rapeseed accessions are suitable for GWAS analysis.

To uncover the relationships of the different yield traits, Pearson correlation coefficient (PCC) was used to assess correlations between pairs of traits in the eight yield-related traits (Figure 1). SN and YP were highly significantly and positively correlated (PCC=0.67), indicating that the YP was greatly determined by the SN. In addition, there were also significant positive correlations between BN and SN, SI and YP, and SI and SN, with their PCCs of 0.47, 0.36 and 0.3, respectively, while SW was negatively correlated with both FT (PCC = –0.36) and BN (PCC = –0.28). These results suggest that there is an intricate relationship between yield-related traits and they play important roles in regulating oilseed rape yield in a coordinated manner.

Genome-wide association mapping for yield traits

A total of 908 QTNs for eight yield traits were identified using at least two of the six ML-GWAS methods, namely FASTmrEMMA, FASTmrMLM, ISIS EM-BLASSO, mrMLM, pKWmEB, and pLARmEB (Supplementary Table S2). QTNs with LOD scores > 3.0 were considered significant trait-related QTNs. The highest number of QTNs for SN was identified to be 127, followed by SI with 126 and the remaining 118, 96, 110, 115, 106, and 110, were associated with PH, FT, BN, SP, SW, and YP, respectively (Supplementary Table S2). The number of QTNs detected by multiple methods varied between environments, with higher numbers found in WH16 and NC14, 435 and 431 respectively, compared to 404 in YL13 (Supplementary Figure S2). The most abundant QTNs for SI were found in both YL13 and WH16, with 76 and 84 each, while the greatest number of QTNs for SN were identified in NC15. FASTmrMLM identified the highest number of

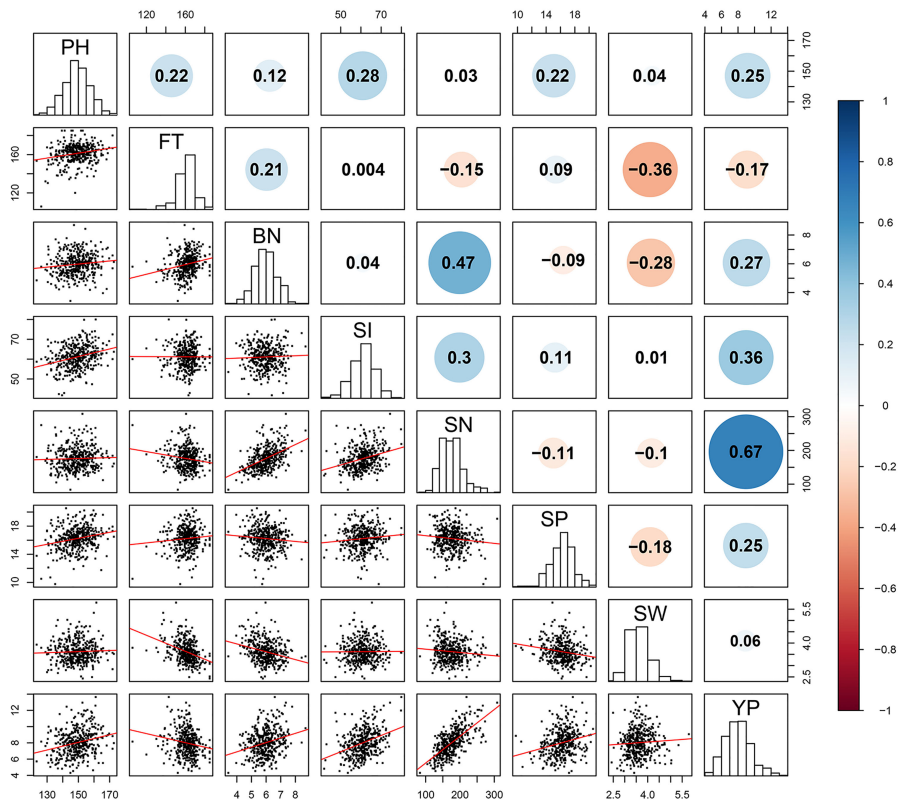


FIGURE 1
Pairwise Pearson correlation among the eight yield traits in *B.napus*. The upper diagonal represents the Pearson correlation coefficient (PCC) between every two traits (positive numbers represent positive correlation, negative numbers represent negative correlation). The diagonal histogram represents the distribution of each trait, and the lower diagonal represents the linear regression statistics between each two traits. Blue for positive correlation, red for negative correlation. The size of the circle represents the absolute magnitude of the PCC.

QTNs (320), while FASTmrEMMA found only 20, with other models identifying numbers in the range of 151 ~ 273 (Supplementary Figure S2). In addition, we found that the results identified by pKWEmEB and pLARmEB were consistent with each other.

We further analyzed the common QTNs that were co-identified in at least four ML-GWAS approaches (Figure 2). Through the combination of different methods, a total of 596 QTNs were identified, of which NC14 was the most, with 236 QTNs, while YL13 was only 154 QTNs. The QTNs associated with SI became the most numerous with 120, the rest being 65, 59, 103, 103, 81, 48, and 65, associated with PH, FT, BN, SN, SP, SW, and YP, respectively (Figure 2). These results showed the diversity of QTNs in different environments or traits, demonstrating the importance of identifying stable QTNs by integration of multiple methods.

Stable QTNs detected by multi-methods or across environments

In order to obtain reliable results, we further analyzed QTNs shared by at least two models in different environments within the 1 Mb region. A total of 75 significant stable QTNs (or QTN clusters) controlling eight yield-related traits were obtained (Table 1 and Supplementary Table S3). One QTN (C03: 12081167) associated with FT in WH16 was identified simultaneously in six models, explaining phenotypic variation of 0.35 ~ 3.79 and LOD scores ranging from 3.04 ~ 5.72, and near which locus, another QTN (C03: 12024986) was also detected in YL13 (Table 1). Six QTNs controlling different traits were detected simultaneously in the five models, with *qSWYL13-A02-1* and *qSWWH16-A02-1* for SW explaining the largest phenotypic variation range of 4.10 ~ 22.24. In addition, all QTNs explained the largest range of phenotypic variation of SW (3.05 ~ 27.33), followed by FT (1.78 ~ 16.40). The number of these QTNs varied considerably in the A and C subgenomes, 23 and 44 respectively, and were more numerous on chromosomes C03, C08, and C09, with 10, 7, and 8, respectively.

Forty QTNs shared by ML-GWAS and SL-GWAS within the 1 Mb region were found to associate with the eight yield traits, with the highest number of QTNs associated with SW (10) and the lowest with SI (2) (Supplementary Table S4) (Hu et al., 2022). Among the overlapped QTNs with SL-GWAS, 10 QTNs including A09: 28130192 and A09:2818207 associated with SW were simultaneously detected, and some of them were found in at least two environments, suggesting that these QTNs were more stable and reliable (Table 2 and Supplementary Table S4). Four QTNs were detected in all five ML-GWAS and SL-GWAS, and two of them, both associated with SW, were found repeatedly in two

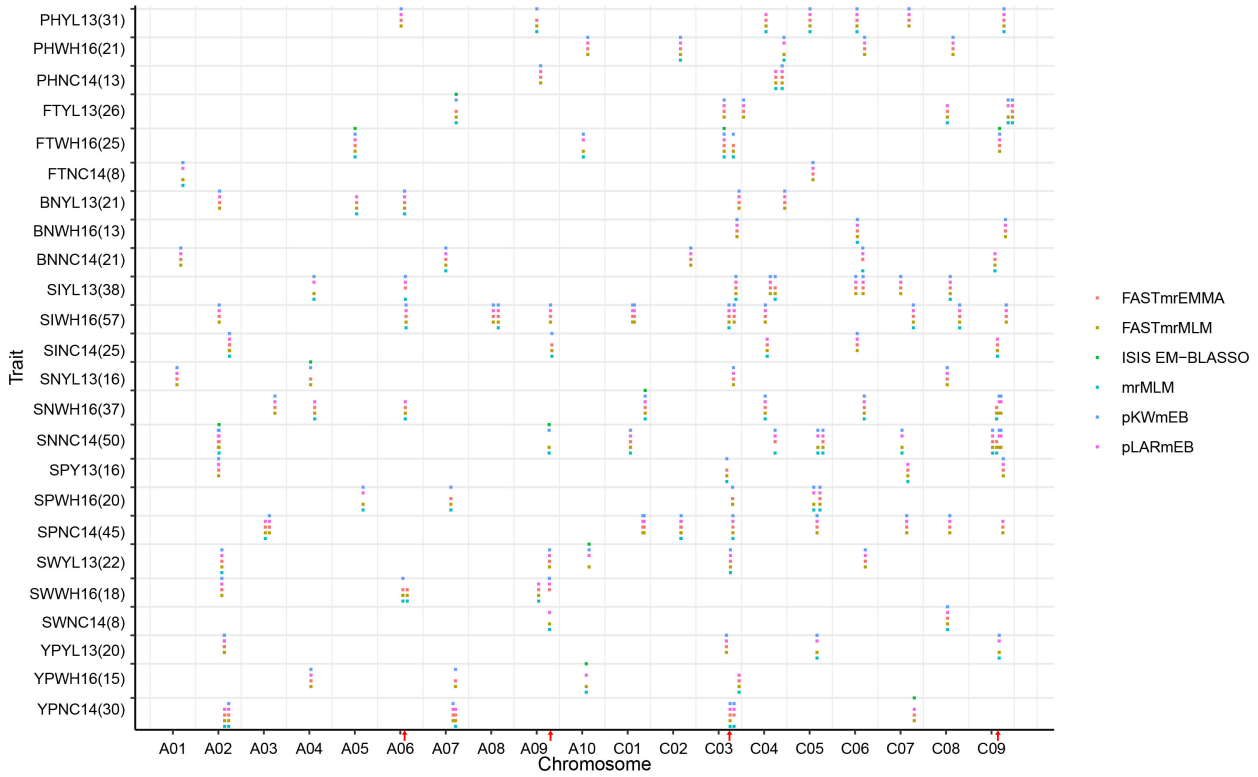


FIGURE 2
Chromosomal distribution of QTNs for eight yield traits identified by six ML-GWAS methods in the three environments. The horizontal axis indicates genomic locations in chromosomal order and plots significant QTNs according to genomic location. Each row represents a QTN identified by a different ML-GWAS methods. PH, plant height; FT, flowering time; BN, primary branch number; SI, Silique number of main inflorescence; SN, Silique number per plant; SP, seed per pod; SW, 1000-seed weight, YP, yield per plant. Three different environments, YL13, Yangluo 2013; NC14, Nanchang 2014; WH16, Wuhan 2016. The red arrows show the QTN hotspots.

10.3389/fpls.2023.1153000

TABLE 1 Candidate genes associated with eight yield-related traits identified by multi loci-GWAS in different environments.

Trait	Chr.	Position	LOD	R^2 (%)	Method	Env.	Candidate genes	Annotation
PH	C05	35061843	5.11-9.57	2.48-2.76	2,3,4	E1	BnaC05g35940D	AUX/IAA protein
	C05	36252249	6.96-14.20	4.18-5.15	2,3,4	E2	BnaC05g36140D	Protein phosphatase 2C
							BnaC05g37110D	AP2/ERF domain
	C08	18928273	4.69-7.07	1.15-2.97	2,3,4	E2	BnaC08g13850D	SANT/Myb domain
	C08	18801051	4.14-6.64	1.77-2.50	1,2,3	E3	BnaC08g13910D	Protein kinase
							BnaC08g13990D	bHLH
FT	A03	23460059	4.76-9.69	2.07-4.25	2,3,5	E1	BnaA03g44520D	Ribonuclease H2
	A03	22616177	6.56-15.01	3.27-4.66	1,2,5	E3	BnaA03g44830D	Pentatricopeptide repeat
							BnaA03g45650D	MADS-box
	C03	12024986	3.07-10.20	2.22-4.77	2,3,4,5	E1	BnaC03g22140D	ABC transporter
	C03	12081167	3.04-5.72	0.35-3.79	1,2,3,4,5,6	E3	BnaC03g22200D	Expansin
							BnaC03g22280D	Amino acid transporter
	C09	45879981	3.77-7.28	1.99-5.74	1,2,3,4,5	E1	BnaC09g45620D	Nonaspanin (TM9SF)
	C09	45879370	4.36-4.83	2.84-4.57	2,4,5	E2	BnaC09g45880D	Cytochrome P450
							BnaC09g45930D	Zinc finger, RING-type
BN	C06	4708955	6.18-8.91	2.11-4.13	2,4,5	E1	BnaC06g03900D	AP2/ERF domain
	C06	4978698	5.22-7.61	2.02-3.11	1,2,3,4,5	E3	BnaC06g04200D	F-box domain, cyclin-like
							BnaC06g04380D	bHLH
	C08	24114714	3.37-8.73	2.17-4.08	2,4,5	E2	BnaC08g21430D	SANT/Myb domain
	C08	25243813	4.31-10.90	1.53-5.04	2,3,4	E3	BnaC08g21500D	Cyclin, C-terminal domain
							BnaC08g22620D	Auxin responsive SAUR
SI	A01	21122950	3.67–7.27	0.51–1.51	2,4,5	E1	BnaA01g31080D	SANT/Myb domain
	A01	21380473	3.07–7.82	0.46–3.18	3,4,5	E3	BnaA01g31290D	AP2/ERF domain
							BnaA01g31770D	F-box domain, cyclin-like
	C02	15059835	4.29–17.47	1.57–6.12	2,4,5	E3	BnaC02g18860D	CWC16 protein
	C02	16019162	3.41–8.08	1.92–3.15	3,4,5	E2	BnaC02g19100D	SBP-box
							BnaC02g19360D	MADS-box
	C03	21828785	3.50–5.00	1.03–2.84	2,3,4	E2	BnaC03g36030D	Armadillo-like helical
	C03	22767136	3.07–6.32	0.27–2.16	1,2,3,4,5	E3	BnaC03g36600D	No apical meristem
							BnaC03g36740D	DNA topoisomerase
SN	A02	13065894	4.29–5.52	1.00–1.89	2,4,5	E1	BnaA02g20660D	Myb/SANT-like domain
	A02	13595586	5.59–8.57	2.44–2.85	2,3,5	E2	BnaA02g20930D	F-box domain, cyclin-like
							BnaA02g21200D	MADS-box
	C06	13960751	3.85–4.74	1.64–2.59	2,4,5	E2	BnaC06g11920D	Protein kinase
	C06	14647381	4.26–7.57	1.76–3.24	2,3,4,5	E3	BnaC06g12120D	Sucrose synthase
	C07	3019188	4.02–6.54	0.92–2.52	2,3,4	E3	BnaC07g02320D	F-box domain, cyclin-like
	C07	3183504	4.18–6.15	1.32–2.43	1,2,4,5	E2	BnaC07g51040D	GH3
							BnaC07g51180D	Aminotransferase
	C09	11236617	6.51–9.86	3.23–3.96	1,2,3	E2,E3	BnaC09g14450D	PI4P5K

(Continued)

10.3389/fpls.2023.1153000

TABLE 1 Continued

Trait	Chr.	Position	LOD	R^2 (%)	Method	Env.	Candidate genes	Annotation
	C09	16839183	4.55–9.55	128–4.06	2,4,5	E2,E3	BnaC09g14650D	Dynamin central domain
	C09	20774003	3.19–5.58	0.88–2.36	2,4,5	E2,E3	BnaC09g14760D	SANT/Myb domain
							BnaC09g19820D	JmjC domain
SP	C03	30605297	5.00–6.14	0.98–2.41	2,3,5	E3	BnaC03g45540D	AGP9
	C03	31339715	4.03–8.30	2.14–4.80	1,2,3,4,5	E2	BnaC03g45610D	GASA7
							BnaC03g46070D	F-box, cyclin-like
							BnaC03g46450D	SAUR protein
	C09	24974617	3.45–6.37	1.45–3.28	2,3,4	E2	BnaC09g25390D	LTPG4
	C09	25836173	3.13–11.90	3.46–5.67	2,3,4,5	E1	BnaC09g25980D	SWEET sugar transporter
							BnaC09g26050D	Peptidase C48, SUMO
							BnaC09g26100D	Protein phosphatase 2C
SW	A02	7610485	7.18–21.14	1.77–8.32	1,2,3,4,5	E1,E3	BnaA02g13530D	SAM methyltransferase
							BnaA02g13870D	Zinc finger, C2H2
							BnaA02g13950D	AP2/ERF domain
	A06	15184188	6.07–14.12	1.68–4.73	1,2,3	E1,E3	BnaA06g21690D	Protein kinase
							BnaA06g21720D	DUF296
							BnaA06g21890D	Zinc finger, RING-CH-type
	A09	27915980	3.85–11.85	2.11–3.22	3,4,5	E3	BnaA09g39450D	Cytochrome b561
	A09	28182807	5.35–9.51	2.44–3.16	1,2,4	E2	BnaA09g39480D	No apical meristem (NAM)
	A09	28130192	4.00–26.19	3.21–7.50	2,3,4,5	E1,E3	BnaA09g39620D	HAD-hydrolase
							BnaA09g39680D	AUX/IAA protein
							BnaA09g39790D	Helicase, C-terminal
							BnaA09g39840D	PMR5 N-terminal domain
							BnaA09g39950D	Lipase, class 3
YP	A07	21399751	5.25–8.52	0.76–4.22	2,3,5	E3	BnaA07g29760D	ABC-2 type transporter
	A07	21594406	4.46–8.12	1.29–1.91	1,2,3,4	E2	BnaA07g30510D	Toll/interleukin-1 receptor
							BnaA07g30950D	Auxin efflux carrier
	A10	8882087	3.20–4.74	0.34–4.04	1,2,4,6	E3	BnaA10g10240D	No apical meristem (NAM)
	A10	9777221	3.93–6.97	2.08–3.50	1,3,4	E1	BnaA10g10870D	Acyl-CoA N-acyltransferase
							BnaA10g11980D	SANT/Myb domain
	C03	44895348	3.45–11.09	1.66–5.47	1,2,3,4	E3	BnaC03g55830D	E3 UFM1-protein ligase 1
	C03	45022247	3.26–4.14	1.06–2.63	2,3,4	E1	BnaC03g55940D	Protein kinase

Chr, Chromosome; Env, Environments; PH, Plant height; FT, Flowering time; BN, Primary branch number; SI, Silique number of main inflorescence; SN, Silique number per plant; SP, Seed per pod; SW, 1000-seed weight, YP, Yield per plant. E1, YL13; E2, NC14; E3, WH16.

environments, explaining 4.10 ~ 22.24, 5.32 ~ 10.60 of the phenotypic variation, respectively (Supplementary Table S4). Notably, we detected two significant signal loci, A09:28182807 and C09:25836173, in three environments simultaneously, associated with SW and SP, respectively. The two loci have high LOD values and may be significant trait-associated QTNs with breeding potential (Supplementary Figure S3). Thus, a total of 24 and 15 stable significantly associated QTNs were detected by ML-GWAS for SW and SP, explaining 0.24 ~ 8.32 and 0.00 ~ 6.74 of phenotypic variation, respectively (Supplementary Table S5).

Since SW is the important component of yield traits, and is relatively stable in different environments in this study, we further analyzed the QTNs associated with SW. A total of 21 significant QTNs were identified simultaneously in at least 2 environments as well as in three models (Table 2). And nine of the 21 QTNs were also detected by SL-GWAS with MLM model (Table 2), indicating

10.3389/fpls.2023.1153000

TABLE 2 Significant 1000-seed weight (SW) associated QTNs and candidate genes detected in at least two environments by multi loci-GWAS.

Chr.	Position	Allele	LOD	R^2(%)	——log10(*P*)	Environment	Methods	MLM (Position)
A02	7610485	T/A	3.39–21.14	4.10–22.24	0.83–8.32	E1,E3	1,2,3,4,5	7600283~7620485
A02	9978238	A/G	3.34–10.95	4.05–11.91	0.89–1.86	E1,E2,E3	1,3,5	9972386~12256474
A04	6184539	T/G	4.18–8.10	4.67–9.00	1.95–2.93	E1,E2,E3	1,2,3	
A05	11305754	C/T	3.30–4.6	4.02–5.43	0.88–2.24	E1,E3	1,2,3,5	
A06	10063511	A/T	7.40–10.03	8.27–10.97	2.43–5.21	E2,E3	1,2,5	
A06	15184188	C/T	4.04–14.12	4.79–15.13	0.82–4.73	E1,E3	1,2,3	
A09	4168712	A/G	3.08–0.34	3.79–11.41	0.99–4.50	E2,E3	1,2,3,4	
A09	28130192	C/T	4.00–26.19	4.75–27.33	1.17–7.50	E1,E2	2,3,4,5	28085608~28158503
A09	28182807	A/G	5.35–17.43	6.16–18.48	2.44–4.26	E1,E2	1,2,4	28171742~28199140
A10	15198168	T/C	3.63–12.37	4.14–13.35	1.11–4.66	E1,E3	2,4,5,6	
C01	11143808	A/C	3.63–7.30	4.36–8.18	1.62–1.92	E1,E2	1,3,5	11143808~15054178
C03	25856914	G/T	4.54–9.67	5.32–10.60	1.94–4.54	E1,E3	1,2,3,4,5	25846914~25866982
C03	33059809	G/A	3.87–8.18	4.62–9.08	0.24–0.65	E1,E3	1,2,3	
C03	38304116	T/G	3.05–7.82	3.05–7.82	1.58–3.51	E1,E2	2,4,5	
C05	6286803	T/C	4.27–5.69	5.03–6.51	1.75–3.62	E1,E2,E3	2,4,5	
C05	30666036	T/A	3.79–6.92	4.53–7.78	1.63–3.23	E1,E3	2,4,5	29675390~30666036
C07	18309538	C/T	3.37–6.54	4.00–7.75	1.04–3.07	E1,E2,E3	2,3,4,5	18306538~22307432
C07	37609075	C/A	5.93–8.90	6.77–9.81	0.88–6.79	E2,E3	1,2,4,5	
C07	38545053	A/G	3.23–10.84	3.23–10.84	0.87–3.68	E1,E3	1,2,3,4	
C08	18157206	C/T	4.50–8.45	5.27–9.35	1.94–5.61	E2,E3	1,2,4	
C08	37165475	A/G	5.33–6.35	6.14–7.19	1.37–4.21	E2,E3	1,2,3,4	31608617~37165765

The six ML-GWAS methods: 1, mrMLM; 2, FASTmrMLM; 3, FASTmrEMMA; 4, pLARmEB; 5, pKWmEB; 6, ISIS EM-BLASSO. E1, YL13; E2, NC14; E3, WH16.

the reliability of ML-GWAS (Hu et al., 2022). A roughly similar number of QTNs was found in each environment, and the results were also similar for each model, except for pLARmEB, which identified only one. These QTNs were randomly distributed on different chromosomes and their LOD scores ranged from 3.05 ~ 26.19, explaining 3.05 ~ 27.33 of the phenotypic variation (Table 2).

Allelic effects of stable QTNs for yield traits

To identify the favorable alleles of QTNs for the eight yield traits in rapeseed, we analyzed the significant phenotypic differences between the elite and alternative alleles of 21 QTNs (Figure 3 and Supplementary Figure S4). We divided the population into two or three groups according to their allele types and compared the phenotypic values among the different groups. Generally, population with elite alleles had significantly larger values than those with unfavorable alleles. For example, accessions with the TT allele of C09:25836173 show more SP compared to those with the AA variant, indicating TT could be considered an elite allele. We focused on A09:28182807, a significant locus associated with SW, and the average SW values of GG individuals of A09:28182807 was significantly higher than that of AA and AG individuals. We further analyzed the haplotype blocks in the 772 bp region around the peak locus A09:28182807 and identified a strongly linked block (Figure 4A). Based on the genotypes of this block, the association population of rapeseed germplasms was divided into five major haplotype groups with two nonsynonymous SNPs (Figure 4B). Haplotype Hap 3 (n =19) had relatively large SW values and showed significant differences from Hap 1 (n = 41) and Hap 5 (n = 18) (Figure 4B). These results suggest that locus A09:28182807 significantly associated with SW and that the haplotype Hap 3 of *BnaA09g39950D* may increase seed weight.

Twelve other QTNs associated with the seven yield-related traits PH, FT, SI, BN, SN, SW and YP were further analyzed, indicating some elite alleles improve the yield traits (Supplementary Figure S4). The favorable alleles obviously affected FT with the population with the AA elite allele at A03:22616177 having a mean FT of around 170 days, compared to around 160 days for the CC allele group (Supplementary Figure S4). These findings suggested that the accessions with elite alleles have clearly higher phenotypic values for yield-related traits compared to those with unfavorable allelic variations. Nineteen important QTNs shared by multiple environments and multiple methods regulate five important yield-related traits, including SW, FT, BN, SI, and SN. And these QTNs were used to assess the utilization of favorable alleles in 20 elite

10.3389/fpls.2023.1153000

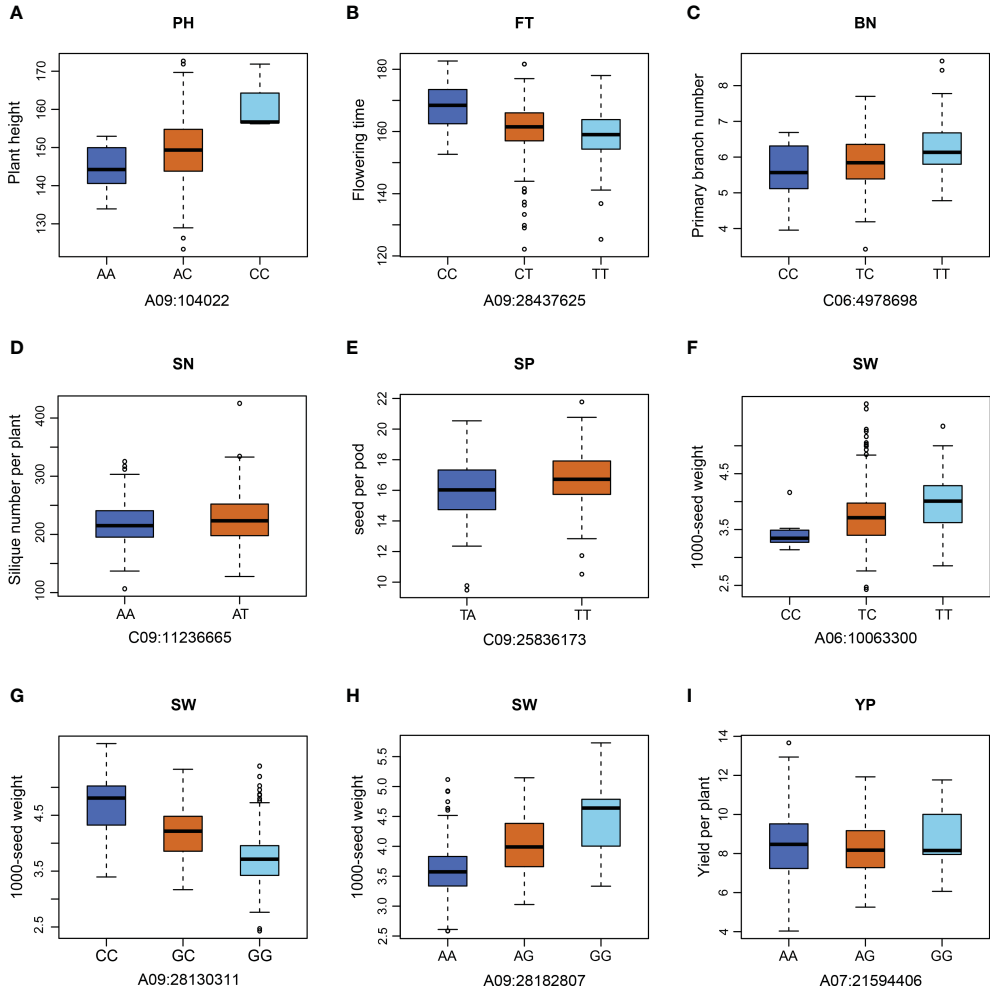


FIGURE 3
Phenotypic differences between two or three genotypes for each of the nine QTNs. **(A-I)** Phenotypic variations at different alleles of nine QTNs for the yield traits. PH, plant height; FT, flowering time; BN, primary branch number; SN, Silique number per plant; SP, seed per pod; SW, 1000-seed weight, YP, yield per plant.

accessions during rapeseed breeding (Supplementary Figure S5). Among these elite accessions, the number of superior alleles in the QTNs ranged from 5 (26.3%, Shengliqinggeng) to 12 (63.2%, Zhongshuang 11), of which seven accessions had more than 10 (52.6%) superior alleles. In addition, we found that some superior alleles of the QTN loci were prevalent in these accessions, for example, the superior alleles of C09:11236665 and C09:11236689 were in almost all the accessions (19/20). These results suggest that some common elite alleles may have a particularly large impact on yield.

Identification of candidate genes based on stable QTNs and transcriptome analysis

Considering the LD decay distance of the rapeseed population, the regions within 200-kb on either side of the stable QTNs based on ML-GWAS were used to identify the candidate genes. Thus, 4796 genes were mined surrounding the 75 stable QTNs identified by ML-GWAS in different environments (Table 1 and Supplementary Tables S4, S5). There are many candidate genes involved in plant growth and development, such as *BnaA07g30950D* (Auxin efflux carrier), which is involved in maintaining the embryonic hormone gradient, and also involved in shoot and root development. The genes *BnaA07g30180D* (SAUR protein), *BnaA09g39680D* (AUX/IAA protein), *BnaA09g40340D* (F-box domain, cyclin-like) and *BnaC03g46450D* (SAUR protein) have all been implicated in the regulation of plant growth (Table 1). Furthermore, 1807 genes were found in the QTNs of SW and 942 genes were discovered in SP (Supplementary Figure S6 and Table S5). Haplotype analysis of the candidate genes *BnaA06g17710D*, *BnaA09g39450D* and *BnaA09g39950D* for SW showed that the different haplotypes had significant phenotypic difference (Supplementary Figure S7). In the significant QTN cluster (A09: 27915980 ~ 28130192 ~ 28182807) for SW, which was also detected by SL-GWAS, three candidate genes *BnaA09g39450D* (Cytochrome b561), *BnaA09g39790D* (RNA helicase), and *BnaA09g39950D* (Lipase) were identified to associate with SW (Figures 3, 4 and Supplementary Table S5).

To further determine the candidate genes associated with the two important traits SW and SP, we analyzed transcriptomic data from two

10.3389/fpls.2023.1153000

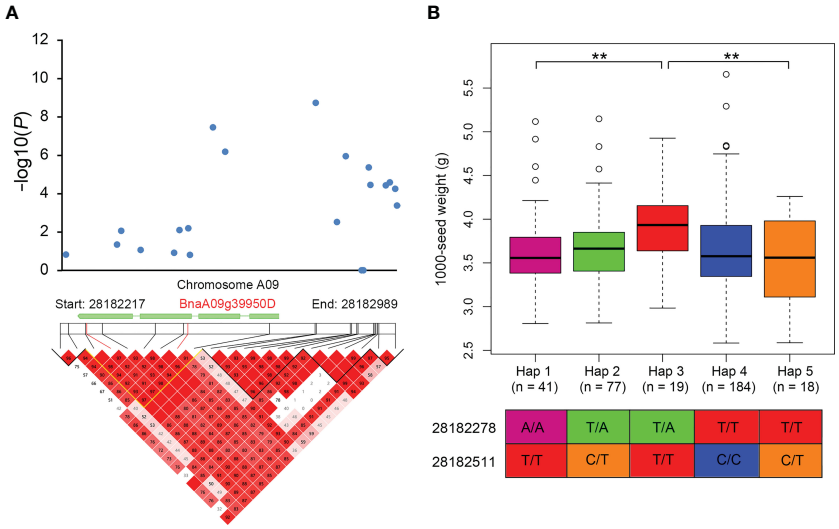


FIGURE 4
The QTN detected in at least two environments on chromosome A09 for 1000-seed weight (SW) with LD heatmap surrounding the QTN. **(A)** Manhattan plot of the A09 chromosomal region around the candidate gene *BnaA09g39950D* and LD heatmap with a peak SNP (A09: 28182807). **(B)** Haplotype analysis of *BnaA09g39950D*. Box plots show the distribution of each haplotype group, n denotes the number of genotypes belonging to each haplotype group and the genotypes less than ten are not shown. ** Significant differences between the haplotypes were evaluated by two-tailed *t* test ($P < 0.01$).

accessions with extremely difference of seed size, R01 and R56. The accession R01 with larger seeds than that of R56, and has significant different seed number per pod (R01, n = 14.74 *vs* R56, n = 20.56) (Figure 5A) (Hu et al., 2022). We found a total of 2572 differentially expressed genes (DEGs) among three different stages in these two accessions (Supplementary Figure S6 and Table S6). Searching for commonly identified genes in the DEGs and the candidate genes for ML-GWAS results, we identified 79 reliable candidates with 50 and 29 of which regulate SW and SP, respectively (Supplementary Figures S6B, C and Tables S7, S8). GO enrichment analysis revealed that DEGs were enriched to fruit development terms (GO: 0010154) in biological processes (GO: 0008150) and identified 76 associated genes (Supplementary Figure S6). KEGG analysis indicated that a large number of DEGs were involved in the metabolic pathways and biosynthesis of secondary metabolism (Figure 5B). Moreover, the gene annotation of DEGs demonstrates considerable genes related to plant hormones, especially auxin, as well as an abundance of transcription factors (TFs).

Analysis of candidate genes expression patterns

Expression pattern analysis was performed for 60 genes detected in both DEGs and ML-GWAS for SW and SP (Figure 5C). We identified a number of genes that were highly significantly differentially expressed in the two cultivars R01 and R56 (Figure 5C). *BnaC09g25980D* (*SWEET7*) is a candidate gene for SP, which is extremely highly expressed in R56 plants and lowly expressed in R01 (Figure 5C). In addition, *BnaA02g05510D*, *BnaA02g05540D*, and *BnaC09g09790D* also had similar expression patterns (Figure 5C). However, *BnaA09g57040D* and *BnaC08g43130D* had opposite expression profiles, both of which were highly up-regulated in R01 and lowly expressed in R56. There are also many genes to be up-regulated in both accessions, but they are more remarkable in R01 than in R56. The SW-related candidate gene *BnaA09g39450D* (Cytochrome b561) was more highly expressed in R01, and expression levels increased progressively with developmental time (Figure 5C).

We next investigated the expression profiles of genes related to different phytohormones, such as auxin, jasmonic acid (JA) and gibberellin (GA), cyclin-related genes, and TFs in the 2572 DEGs (Supplementary Figures S8A–E). Transcripts of seven auxin-related genes were up-regulated in R56 compared to R01, for example, *BnaA05g00250D*, *BnaC04g00140D* and *BnaC09g00360D* had higher expression levels in R56 (Supplementary Figure S8A). All three genes related to JA were extremely up-regulated in R56, but *BnaA02g05120D* was expressed at a higher level in R01 (Supplementary Figure S8B). Three genes related to, GAs especially *BnaC03g11560D* were also up-regulated in R56, but two cyclin genes, *BnaC01g38940D,* and *BnaC02g02720D*, were significantly higher expressed in R01 than in R56 (Supplementary Figures S8C, D). In addition, 36 TFs in 2572 DEGs were showed differentially expressed in R01 and R56 (Supplementary Figure S8E).

Validation of DEGs by qRT-PCRs

Twelve candidate genes were selected for qRT-PCR analysis based on the combined results of ML-GWAS and RNA-seq (Figure 6). Overall, the expression levels of these genes were diverse in the two accessions and varied across developmental periods. The expression profile of the candidate gene *BnaC09g25980D* for SP was consistent with RNA-seq, with both showing high expression levels of in R56 and lowly expression in R01. The qRT-PCR validation of the

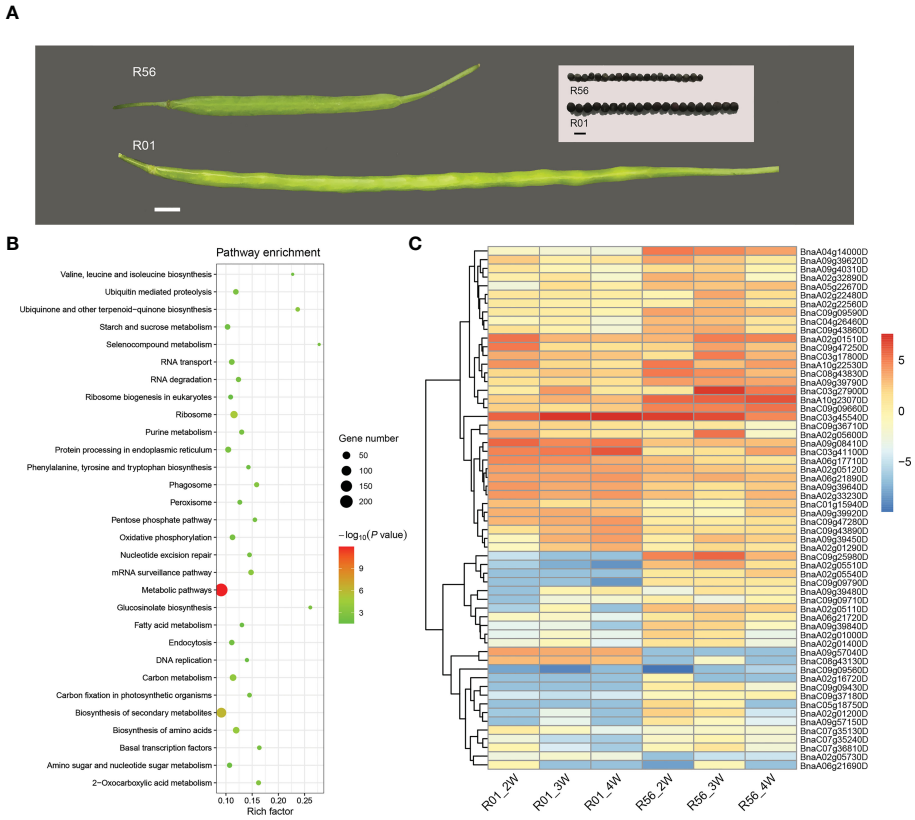


FIGURE 5
Transcriptome analysis of two rapeseed accessions with extremely seed size/seed weight. **(A)** Seeds and siliques of R01 (large seed) and R56 (small seed) showed considerable variations in seed weight. **(B)** The top 30 significantly enriched KEGG pathways of differentially expressed genes (DEGs) between R01 and R56 at three stages. **(C)** Heatmap of the expression patterns of the 60 genes in developing seeds of two rapeseed cultivars at three stages. The red indicates high expression, and the blue shows low expression. R01_2W, R01_3W, R01_4W, R56_2W, R56_3W and R56_4W represent the sampling time of R01 and R56 are two weeks after pollination, three weeks after pollination, and four weeks after pollination, respectively.

candidate gene *BnaA09g39450D* at the significant signal locus A09:28182807 for SW showed that it was expressed in both accessions, but transcripts were more abundant in R01 and increased progressively over time. Interestingly, the genes *BnaC09g43860D* and *BnaC04g26460D* were lowly expressed in R01 and expressed at high levels in R56, but gradually decreased in R56 over time. There are three genes, such as *BnaA06g21890D*, *BnaC03g45540D* and *BnaA10g23070D*, which are expressed in both accessions and their expression levels increase progressively with development in R01 but decrease progressively in R56. For gene *BnaA02g05540D*, it was extremely lowly expressed at different developmental periods in both R01 and R56, except for R01 at 4WAP. In addition, we found that most of these genes were significantly differentially expressed in both accessions, such as *BnaA06g17710D*, *BnaA09g08410D*, *BnaC09g43890D*, and *BnaC09g47280D*.

Discussion

Seed yield is an important and complex quantitative trait in rapeseed. And developing high yield rapeseed varieties is the goal for breeders in many cases. Previous studies did not comprehensive analysis of the seed yield and yield-related traits in different environments at the same time using whole genome resequencing. In the present study, we not only examined the relationship among eight yield traits, but also detected novel QTNs for the yield traits in three environments. These findings would be helpful for breeders to develop rapeseed varieties by congregating superior alleles.

In rapeseed, lots of complex traits have been dissected *via* the GLM or MLM based on the single-locus using arrays or resequencing data (Lu et al., 2019; Song et al., 2020; Wang et al., 2021a; Hu et al., 2022). However, using one model for GWAS has limitations and may miss the small-effect loci (Ma et al., 2018). In this study, we use multi-locus methodologies, including mrMLM, FASTmrEMMA, pLARmEB and so on, to detect novel and more loci for seed yield and yield-related traits in rapeseed. Using ML-GWAS, a total of 908 QTNs were identified by at least two of six ML-GWAS methods and 596 QTNs of them were obtained for integrating multiple approaches (Figure 2 and Supplementary Table S2). And 40 loci were the same and near the loci which have been detected in rapeseed using SL-GWAS with MLM model as previous study (Hu et al., 2022). Furthermore, 75 of new QTNs have been identified in different environments and at least two

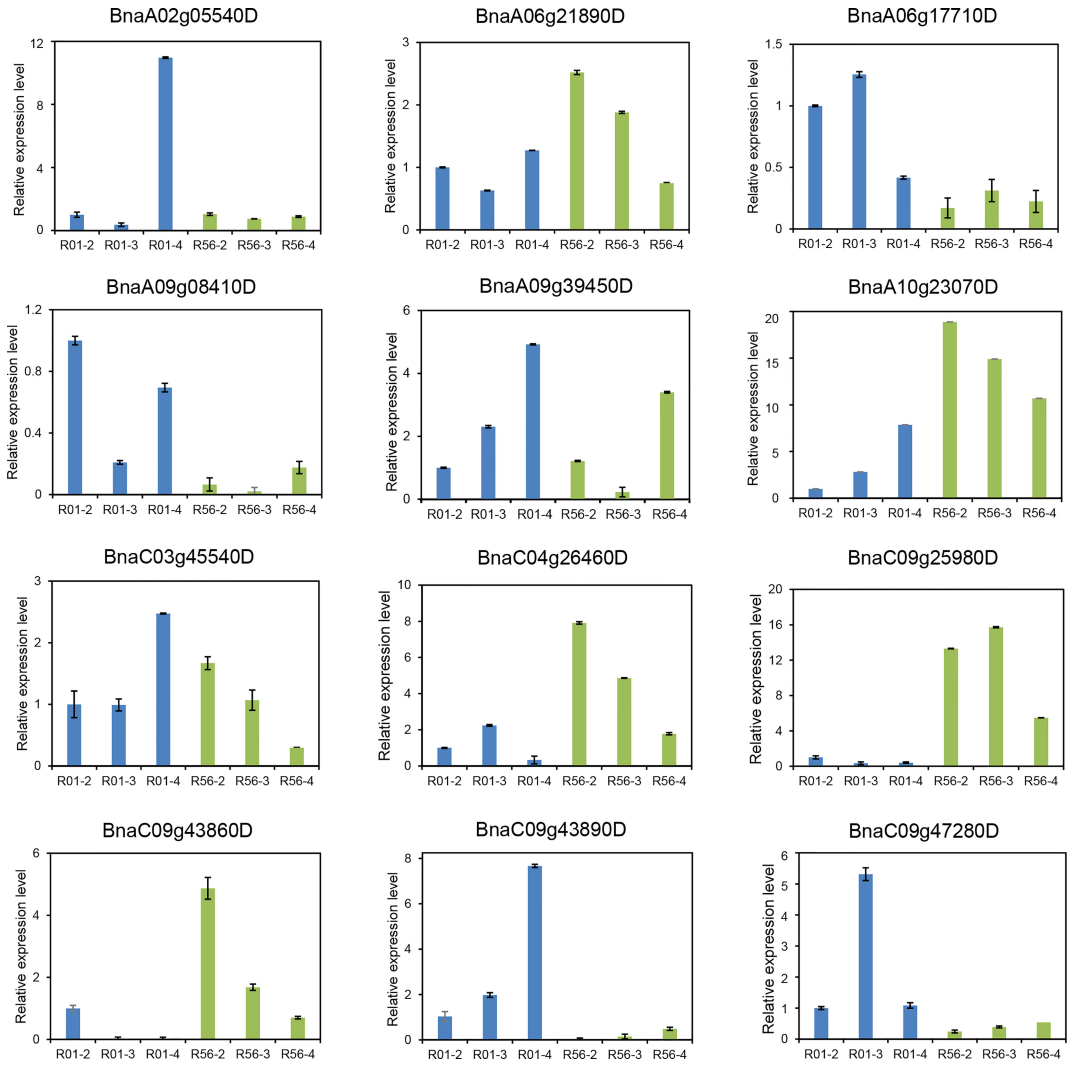


FIGURE 6
Quantitative RT-PCR (qRT-PCR) validation of differentially expressed genes (DEGs) for 1000-seed-weight (SW) between R01 and R56 from RNA-seq data. The transcript abundances were calculated from three replicates with *BnACTIN7* (*BnaA03g55890D*) as internal control. Data are shown as means ± SE. R01-2, R01-3, R01-4, R56-2, R56-3 and R56-4 represent the sampling time of R01 and R56 are two weeks after pollination, three weeks after pollination, and four weeks after pollination, respectively.

models (Supplementary Table S3). Meanwhile, each method successfully detected some loci which other methods were not identified, indicating that it is worth using various methods for GWAS analysis (Liu et al., 2020). The stable QTNs identified in two or more environments or at least two methods increased the reliability of these loci. Using these methods, many novel loci were discovered for the seed yield and yield related traits of rapeseed (Supplementary Tables S2-S4). Thus, these ML-GWAS approaches would be effective alternative methods to dissect the genetic architecture for agronomic traits. In addition, integration of GWAS and transcriptomic analysis could reliably identify candidate genes for seed yield in rapeseed (Lu et al., 2017; Zhang et al., 2022a).

Seed size, an important agronomic trait determining crop yield, affects the SW and SP in rapeseed. In the significant QTN cluster on chromosome A09 for SW, three candidate genes *BnaA09g39450D* (Cytochrome b561), *BnaA09g39790D* (RNA helicase), and *BnaA09g39950D* (Lipase) were identified to affect seed size (Figures 3, 4 and Supplementary Table S4). Cytochrome b561, which plays an important role in plant growth and development has been also identified the QTL *qGY8.1* for yield in rice (Balakrishnan et al., 2020). In our study, the expression level of *BnaA09g39450D* was validated to be higher in R01 than that in R56 (Figure 6). In *Arabidopsis*, RNA helicase has been reported to participate in coordination between cell cycle progression and cell size, which is required for ovule development and involved in seed size regulation (Yoine et al., 2006; Bush et al., 2015). Lipase is reported to involve in seed oil production in many plants (Eastmond, 2006; Wang et al., 2021b). In most cases, the seed oil content positively correlated with seed size in *B. napus*. For example, six nonspecific phospholipase C

(NPC) genes have been identified to associate with SW or YP and the favorable haplotype of *BnNPC6.C01* could increase seed oil content and seed yield (Cai et al., 2020). Patatin-related phospholipase pPLAIIIδ has been reported to affect organ size (silique) in *Arabidopsis* and *B. napus* (Dong et al., 2014). Therefore, these three candidate genes might be involved in the regulation of seed weight and seed yield in rapeseed. Seed size is also determined by the carbohydrate *via* the phloem to developing seeds during seed filling (Sosso et al., 2015). In soybean, both GmSWEET10a and GmSWEET10b were shown to transport sucrose and hexose, contributing to sugar allocation, which consequently simultaneous increases oil content and seed size in soybean (Wang et al., 2020). In our study, the candidate gene *BnaC09g25980D* (*SWEET7*) was validated to be highly expressed in R56 (Figure 6), indicating that *SWEET7* might regulate the seed size in rapeseed.

Many plant hormones have been reported to be involved in the regulation of seed size, including auxin pathway, GA signaling and brassinosteroid (BR) signaling (Li et al., 2019c). In this study, using the extremely difference of seed size accessions R01 and R56 with significantly different seed number per pod, 2572 DEGs were identified by RNA-seq (Figure 5 and Supplementary Figure S6). And we identified 58 DEGs related to phytohormones, cell cycle and TFs including NAC, TCP, MYB and so on (Supplementary Figure S8 and Table S5). Auxin is known to regulate plant growth and development *via* cell division and cell elongation (Hu et al., 2021b). In previous studies, auxin signaling genes *ARF18* and *BnaA3.IAA7* have been reported to regulate seed weight and yield in *B. napus* (Liu et al., 2015; Li et al., 2019b). In our study, several candidate genes and ten DEGs were found to involve in auxin signaling pathway (Table 1, Supplementary Figure S8 and Table S6). The GASA family in *Arabidopsis* is regulated by GA, with the GASA4 mutant having smaller seeds than the wild type and increased grain weight after overexpression (Roxrud et al., 2007). A number of GA-related candidate genes, including *GASA10* (*BnaC03g11560D*) was highly expressed in R56, which may affect the seed size (Supplementary Figure 8C). Hu et al. (2021c) revealed that JA signaling represses seed size and negatively regulates cell proliferation of integument during seed development. The JA signaling repressor *jaz6* mutants in *Arabidopsis* exhibited small seed size, and overexpression of *BnC08.JAZ1-1* in *Arabidopsis* resulted in enhanced seed weight (Hu et al., 2021c; Wang et al., 2022). In this study, *JAZ12* (*BnaA02g05120D*) and *JAZ 9* (*BnaA07g28810D*) were up-regulated in R01, while another candidate gene *JAZ10* (*BnaC09g43860D*) was highly expressed in R56 (Figure 6, Supplementary Figure 8B and Table S6). These results suggested that auxin, GA, and JA signaling genes were involved in the control of seed size and seed number per pod in rapeseed. Furthermore, we observed that several cell cycle genes, including *BnaC04g26460D* (*CDKB1;1*) were up-regulated in R56 (Figure 6 and Supplementary Figure S8), indicating these DEGs play important roles in cell division and seed size (Qi et al., 2012; Hu et al., 2021c). In addition, one of the candidate gene *BnaC09g43890D* (*NAC083*) was validated highly expressed in R01 (Figure 6). In rice, three NAC genes *NAC020*, *NAC026* and *NAC023* have been reported to associate with seed size/weight (Mathew et al., 2016). And *VvNAC26* was also demonstrated to regulate the seed size by interacting with VvMADS9 in grapevine (Zhang et al., 2021). Therefore, further functional studies of these genes associated with yield traits will help to elucidate the mechanism of high yield and apply to develop high-yielding rapeseed varieties.

Conclusion

In this study, a total of 908 QTNs were detected for eight yield traits using two or more ML-GWAS methods. Of them, 75 stable QTNs (or QTN clusters) controlling yield traits were obtained with a significant QTN cluster on chromosome A09 for SW, which was also identified by SL-GWAS. Twenty elite rapeseed accessions had a diverse distribution of superior alleles, and the high-yielding accessions contained more superior alleles. Integrated ML-GWAS with transcriptome analysis, 79 candidate genes were found to associate with SW or SP. Some genes related to plant hormones such as auxin, JA, and GA, were involved in the regulation of rapeseed yield. Thus, many robust QTLs with candidate genes were identified to regulate seed size and yield traits in rapeseed. This study made a beneficial attempt *via* a combinatory approach of ML-GWAS methods to facilitate the detection of yield-related QTNs in rapeseed. These findings will provide valuable information for understanding the mechanism underlying seed yield and yield-related traits and accelerate the crop improvement of 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

JH and SL designed and supervised the research. SL, HZ, JH, YL and CD performed multi-locus GWAS and bioinformatics analysis. CZ, RG, RZ, JD and JH performed the experiments and data analysis. CZ, RZ, JH and HZ wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The work was supported by the National Natural Science Foundation of China (31901426) and the Key R&D Program of Shaanxi Province (2023-YBNY-031).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1153000/full#supplementary-material

References

Balakrishnan, D., Surapaneni, M., Yadavalli, V. R., Addanki, K. R., Mesapogu, S., Beerelli, K., et al. (2020). Detecting CSSLs and yield QTLs with additive, epistatic and QTL×environment interaction effects from *Oryza sativa* × *O. nivara* IRGC81832 cross. *Sci. Rep.* 10, 7766. doi: 10.1038/s41598-020-64300-0

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21, 263–265. doi: 10.1093/bioinformatics/bth457

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 23 (19), 2633–2635. doi: 10.1093/bioinformatics/btm308

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325. doi: 10.1093/nar/gkab447

Bush, M., Crowe, N., Zheng, T., and Doonan, J. (2015). The RNA helicase, elF4A-1, is required for ovule development and cell size homeostasis in *Aradidopsis*. *Plant J.* 84, 989–1004. doi: 10.1111/tpj.13062

Cai, G., Fan, C., Liu, S., Yang, Q., Liu, D., Wu, J., et al. (2020). Nonspecific phospholipase C6 increase seed oil production in oilseed brassicaceae plants. *New Phytol.* 226, 1055–1073. doi: 10.1111/nph.16473

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

Chen, B., Xu, K., Li, J., Li, F., Qiao, J., Li, H., et al. (2014). Evaluation of yield and agronomic traits and their genetic variation in 488 global collections of *Brassica napus l. Genet. Resour Crop Evol.* 61 (5), 979–999. doi: 10.1007/s10722-014-0091-8

Cui, Y., Zhang, F., and Zhou, Y. (2018). The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01464

Dong, Y., Li, M., Zhang, P., Wang, X., Fan, C., and Zhou, Y. (2014). Patatin-related phospholipase pPLAIIIδ influences auxin-responsive cell morphology and organ size in *Arabidopsis* and *Brassica napus*. *BMC Plant Biol.* 14, 332. doi: 10.1186/s12870-014-0332-1

Eastmond, P. (2006). SUGAR-DEPENDENT1 encodes a patatin domain triacylglycerol lipase that initiates storage oil breakdown in germinating *Arabidopsis* seeds. *Plant Cell.* 18, 665–675. doi: 10.1105/tpc.105.040543

He, Y., Wu, D., Wei, D., Fu, Y., Cui, Y., Dong, H., et al. (2017). GWAS, QTL mapping and gene expression analyses in *Brassica napus* reveal genetic control of branching morphogenesis. *Sci. Rep.* 7 (5), 15971–15976. doi: 10.1038/s41598-017-15976-4

Hu, J. H., Chen, B. Y., Zhao, J., Zhang, F. G., Xie, T., Xu, K., et al. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. *Nat. Genet.* 54 (5), 694–704. doi: 10.1038/s41588-022-01055-6

Hu, J. H., Huang, L., Chen, G., Liu, H., Zhang, Y., Zhang, R., et al. (2021b). The elite alleles of *OsSPL4* regulate grain size and increase grain yield in rice. *Rice* 14, 90. doi: 10.1186/s12284-021-00531-7

Hu, S., Yang, H., Gao, H., Yan, J., and Xie, D. (2021c). Control of seed size by jasmonate. *Sci. China Life Sci.* 64, 1215–1226. doi: 10.1007/s11427-020-1899-8

Hu, J. H., Zhang, F. L. G., Gao, G. Z., Li, H., and Wu, X. M. (2021a). Auxin-related genes associated with leaf petiole angle at the seedling stage are involved in adaptation to low temperature in *Brassica napus*. *Environ. Exp. Bot.* 182, 104308. doi: 10.1016/j.envexpbot.2020.104308

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Li, F., Chen, B., Xu, K., Gao, G., Yan, G., Qiao, J., et al. (2016). A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*). *Plant Sci.* 242, 169–177. doi: 10.1016/j.plantsci.2015.05.012

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 (4), 274–2760. doi: 10.1104/pp.15.01040

Li, H., Li, J., Song, J., Zhao, B., Guo, C., Wang, B., et al. (2019b). An auxin signaling gene *BnaA3.IAA7* contributes to improved plant architecture and yield heterosis in rapeseed. *New Phytol.* 222 (2), 837–851. doi: 10.1111/nph.15632

Li, N., Song, D., Peng, W., Zhan, J., Shi, J., Wang, X., et al. (2019a). Maternal control of seed weight in rapeseed (*Brassica napus* l.): the causal link between the size of the pod (mother, source) and seed (offspring, sink). *Plant Biotechnol. J.* 17 (4), 736–749. doi: 10.1111/pbi.13011

Li, N., Xu, R., and Li, Y. (2019c). Molecular networks of seed size control in plants. *Ann. Rev. Plant Biol.* 70, 435–463. doi: 10.1146/annurev-arplant-050718-095851

Lin, Y., Zhou, K., Hu, H., Jiang, X., Yu, S., Wang, Q., et al. (2021). Multi-locus genome-wide association study of four yield-related traits in Chinese wheat landraces. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.665122

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–E5132. doi: 10.1073/pnas.1502160112

Liu, S., Zhong, H., Meng, X., Sun, T., Li, Y., Pinson, S., et al. (2020). Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. *BMC Plant Biol.* 20 (1), 441. doi: 10.1186/s12870-020-02603-0

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-delta delta c) method. *Methods.* 25 (4), 402–408. doi: 10.1006/METH.2001.1262

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, 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. doi: 10.3389/fpls.2017.00206

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* 10 (1), 1154-1163. doi: 10.1038/s41467-019-09134-9

Ma, L., Liu, M., Yan, Y., Qing, C., Zhang, X., Zhang, Y., et al. (2018). Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-w de association studies. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00561

Mathew, I., Das, S., Mahto, A., and Agarwal, P. (2016). Three rice NAC transcription factors heteromerize and are associated with seed size. *Front. Plant Sci.* 7, 1638. doi: 10.3389/fpls.2016.01638

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 (9), 1297–1303. doi: 10.1101/gr.107524.110

Misra, G., Badoni, S., Anacleto, R., Graner, A., Alexandrov, N., and Sreenivasulu, N. (2017). Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. *Sci. Rep.* 7 (1), 12478. doi: 10.1038/s41598-017-12778-6

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Qi, P., Lin, Y., Song, X., Shen, J., Huang, W., Shan, J., et al. (2012). The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating cyclin-T1;3. *Cell Res.* 22, 1666–1680. doi: 10.1038/cr.2012.151

Raboanatahiry, N., Chao, H., Dalin, H., Pu, S., Yan, W., Yu, L., et al. (2018). Alignment for seed yield and yield related traits in *Brassica napus*. *Front. Plant Sci.* 9, 1127. doi: 10.3389/fpls.2018.01127

Roxrud, I., Lid, S. E., Fletcher, J. C., Schmidt, E. D., and Opsahl-Sorteberg, H. G. (2007). GASA4, one of the 14-member *Arabidopsis* GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol.* 48 (3), 471–483. doi: 10.1093/pcp/pcm016

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 (3), 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 (3), 524–539. doi: 10.1111/tpj.14236

Song, J., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and WWecotype differentiation of *Brassica napus*. *Nat. Plants* 6 (1), 34–45. doi: 10.1038/s41477-019-0577-7

Sosso, D., Luo, D., Li, Q. B., Sasse, J., Yang, J., Gendrot, G., et al. (2015). Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat. Genet.* 47, 1489–1493. doi: 10.1038/ng.3422

Tian, T., Liu, Y., Yan, H., 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

Wang, J., Fan, Y., Mao, L., Qu, C., Lu, K., Li, J., et al. (2021a). Genome-wide association study and transcriptome analysis dissect the genetic control of silique length in *Brassica napus l. Biotechnol. Biofuels.* 14 (1), 214. doi: 10.1186/s13068-021-02064-z

Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016). Improving power and accuracy of genome-wide association studies *via* a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 19444. doi: 10.1038/srep19444

Wang, Y., Li, N., Zhan, J., Wang, X., Zhou, X., Shi, J., et al. (2022). Genome-wide analysis of the JAZ subfamily of transcription factors and functional verification of *BnC08.JAZ1-1* in *Brassica napus. Biotechnol. Biofuels Bioprod.* 15, 93. doi: 10.1186/s13068-022-02192-0

Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *Natl. Sci. Rev.* 7 (11), 1776–1786. doi: 10.1093/nsr/nwaa110

Wang, H., Wang, Q., Pak, H., Yan, T., Chen, M., Chen, X., et al. (2021b). Genome-wide association study reveals a patatin-like lipase relating to the reduction of seed oil content in *Brassica napus*. *BMC Plant Biol.* 21, 6. doi: 10.1186/s12870-020-02774-w

Yoine, M., Nishii, T., and Nakamura, K. (2006). *Arabidopsis* UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. *Plant Cell Physiol.* 47 (5), 572–580. doi: 10.1093/pcp/pcj035

Zhang, S., Dong, R., Wang, Y., Li, X., Ji, M., and Wang, X. (2021). NAC domain gene VvNAC26 interacts with VvMADS9 and influences seed and fruit development. *Plant Physiol. Biochem.* 164, 63–72. doi: 10.1016/j.plaphy.2021.04.031

Zhang, Y. M., Jia, Z., and Dunwell, J. M. (2019). Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00100

Zhang, Y., Li, P., Zhang, J., Li, Y., Xu, A., and Huang, Z. (2022b). Genome-wide association studies of salt tolerance at the seed germination stage and yield-related traits in *Brassica napus* l. *Int. J. Mol. Sci.* 23, 15892. doi: 10.3390/ijms232415892

Zhang, K., Liu, S., Li, W., Liu, S., Li, X., Fang, Y., et al. (2018). Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01690

Zhang, Y. W., Tamba, C. L., Wen, Y. J., Li, P., Ren, W. L., Ni, Y. L., et al. (2020). mrMLM v4.0.2: An r platform for multi-locus genome-wide association studies. *Genomics Proteomics Bioinf.* 18 (4), 481–487. doi: 10.1016/j.gpb.2020.06.006

Zhang, R., Zhang, C., Yu, C., Dong, J., and Hu, J. H. (2022a). Integration of multi-omics technologies for crop improvement: Status and prospects. *Front. Bioinform.* 19. doi: 10.3389/fbinf.2022.1027457

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

Zhong, H., Liu, S., Meng, X., Sun, T., Deng, X., Peng, Z., et al. (2021b). Uncovering the genetic mechanisms regulating panicle architecture in rice with GPWAS and GWAS. *BMC Genomics* 22, 86. doi: 10.1186/s12864-021-07391-x

Zhong, H., Liu, S., Sun, T., Kong, W., Deng, X., Peng, Z., et al. (2021c). Mulit-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* 21, 364. doi: 10.1186/s12870-021-03146-8

Zhong, H., Liu, S., Zhao, G., Zhang, C., Peng, Z., Wang, Z., et al. (2021a). Genetic diversity relationship between grain quality and appearance in rice. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.708996

Zhou, Q. H., Fu, D. H., Mason, A. S., Zeng, Y. J., Zhao, C. X., and Huang, Y. J. (2014). In silico integration of quantitative trait loci for seed yield and yield-related traits in *Brassica napus*. *Mol. Breeding.* 33 (4), 881–894. doi: 10.1007/s11032-013-0002-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fpls.2023.1132302

OPEN ACCESS

EDITED BY
Xiangshu Dong,
Yunnan University, China

REVIEWED BY
Zhansheng Li,
Insititute of Vegetables and Flowers (CAAS), China
Prabhakaran Soundararajan,
National Institute of Plant Genome Research (NIPGR), India
Franziska S. Hanschen,
Leibniz Institute of Vegetable and Ornamental Crops, Germany

*CORRESPONDENCE
Tomohiro Kakizaki
tkaki@affrc.go.jp

RECEIVED 27 December 2022
ACCEPTED 05 April 2023
PUBLISHED 06 June 2023

CITATION
Endo R, Chikano H, Itabashi E, Kawasaki M, Ohara T and Kakizaki T (2023) Large insertion in radish *GRS1* enhances glucoraphanin content in intergeneric hybrids, *Raphanobrassica* (*Raphanus sativus* L. x *Brassica oleracea* var. *acephala*). *Front. Plant Sci.* 14:1132302. doi: 10.3389/fpls.2023.1132302

Large insertion in radish *GRS1* enhances glucoraphanin content in intergeneric hybrids, *Raphanobrassica* (*Raphanus sativus* L. x *Brassica oleracea* var. *acephala*)

Ryota Endo[1], Hiroshi Chikano[1], Etsuko Itabashi[2], Mitsuyo Kawasaki[2], Takayoshi Ohara[2] and Tomohiro Kakizaki[2]*

[1]Agricultural and Bio Resource Development Department, Innovation Division, KAGOME CO., LTD., Nasushiobara, Japan, [2]Institute of Vegetable and Floriculture Science, National Agriculture and Food Research Organization, Tsu, Japan

Glucosinolates (GSLs), precursors of isothiocyanates (ITCs), are present in *Brassicaceae* plants have been found to have health benefits. Sulforaphane (4-(methylsulfinyl)butyl ITC) is an ITC stored in the form of 4-(methylsulfinyl)butyl GSL (glucoraphanin, 4MSOB) in *Brassica* vegetables, such as broccoli and kale. Sulforaphane activates Nrf2 expression, a transcription factor responsible for inducing physiological activities such as detoxification in the human body, and it represents a functional component unique to cruciferous vegetables. *Raphanobrassica* is an inter-generic hybrid between radish and kale, and it contains a high amount of 4MSOB. However, *Raphanobrassica* contains as much 4-methylsulfinyl-3-butenyl GSL (glucoraphenin, 4MSO3B) as it does 4MSOB. GLUCORAPHASATIN SYNTHASE 1 (GRS1) is an enzyme present in radish that synthesizes 4-methylthio-3-butenyl GSL (glucoraphasatin, 4MT3B), a precursor of 4MSO3B, using 4-(methylthio)butyl GSL (glucoerucin, 4MTB) as a substrate. Since the precursor of 4MSOB is also 4MTB, it was considered that both 4MSOB and 4MSO3B accumulate owing to competition in *Raphanobrassica*. We hypothesized that owing to the impaired function of GRS1 in *Raphanobrassica*, it may be possible to breed *Raphanobrassica* cultivars containing a high 4MSOB content. In this study, we generated *Raphanobrassica* populations with functional and defective *GRS1* and compared the GSL composition in the two populations using high-performance liquid chromatography. The mean 4MSOB content in leaves of the defective-type populations was higher than that in the functional-type population, and the defective/functional ratio ranged from 2.02 to 2.51-fold, supporting this hypothesis. Furthermore, leaves, flower buds, stems, and roots contained higher amounts of 4MSOB in the defective population than in the functional population. The leaf 4MSOB content of defective *Raphanobrassica*

grown in this study was comparable to that of previously studied vegetables (such as broccoli sprouts) with high 4MSOB content. *Raphanobrassica* with defective *GRS1* represents a new leafy vegetable with high 4MSOB content which exhibits anti-cancerous and anti-inflammatory potentials.

KEYWORDS

glucosinolate, radish, kale, intergeneric hybrid, *Raphanobrassica*, *GLUCORAPHASATIN SYNTHASE 1*, sulforaphane, glucoraphanin

1 Introduction

More than 500 plant species, primarily cruciferous plants, have been found to contain glucosinolates (GSLs). The common structure of GSLs consists of a β-D-glucopyranose residue linked by a sulfur atom to a (Z)-*N*-hydroximinosulfate ester, plus a variable side chain derived from precursor amino acids (Fahey et al., 2001; Halkier and Gershenzon, 2006). Based on the type of amino acid from which they are produced, GSLs can be categorized as aliphatic, benzenic, and indolic GSLs. Methionine (or alanine, isoleucine, leucine, and valine as alternatives), phenylalanine, and tryptophan are precursors of aliphatic, benzenic, and indolic GSLs, respectively. When plant tissues are damaged by herbivores and/or during infection, GSLs are hydrolyzed by myrosinases to isothiocyanates, thiocyanates, nitriles, or epithionitriles, depending on pH and the presence of epithiospecifier protein (Bones and Rossiter, 1996; Rask et al., 2000).

4-(Methylsulfinyl)butyl GSL (glucoraphanin, 4MSOB), which is a precursor of sulforaphane, is an aliphatic GSL. Sulforaphane has several physiological activities, one of which involves the activation of the transcription factor NF-E2–related factor 2 (Nrf2) (Kensler et al., 2013; Yuanfeng et al., 2021). Nrf2 is a master regulator of detoxification and antioxidants, and it controls the expression of downstream antioxidant genes and phase II detoxification enzyme genes by activating the oxidation response element (Chen and Maltagliati, 2018). The health benefits of sulforaphane have been widely studied in humans; sulforaphane intake reduces the level of urinary 8-hydroxyguanosine, an oxidative stress marker, and it decreases gamma-glutamyl transpeptidase and alanine transaminase levels, which are indicators of liver dysfunction (Kikuchi et al., 2015). Cognitive function is improved by the simultaneous intervention of sulforaphane intake and brain training (Nouchi et al., 2021). In addition, sulforaphane intake mediates the excretion of mycotoxins and air pollutants and improves mild asthma symptoms (Kensler et al., 2005; Riedl et al., 2009; Egner et al., 2014).

4MSOB accumulates in certain *Brassica oleracea* vegetables (2n = 2x = 18, CC genome; Figure 1A) (Cartea and Velasco, 2007). Broccoli (*B. oleracea* var. *italica*) contains the highest amount of 4MSOB among the currently evaluated *Brassica* vegetables. Farnham et al., (2004) reported that the average of 4MSOB concentration in 32 broccoli variety was 0.36 μmol·g^{-1} fresh weight and the range of 4MSOB was 0.24 to 1.85 μmol·g^{-1} fresh weight. It is known that 4MSOB accumulates in all tissues of broccoli and its content is particularly high in mature seeds and seedlings (Yagishita et al., 2019). 4MSOB is also known to be contained in red cabbage (Wermter et al., 2020). Several studies have attempted to increase the 4MSOB content in broccoli, and high 4MSOB-containing broccoli has been bred from crosses between a closely related wild species, *Brassica villosa*, and broccoli (Traka et al., 2013). Recently, it was suggested that MYB28 derived from *B. villosa* increases the transcription of genes that encode GSL biosynthesis enzymes in broccoli (Neequaye et al., 2022). Another 4MSOB-rich vegetable is allopolyploid *Raphanobrassica* (2n = 4x = 36, RRCC), an intergeneric hybrid of the genera *Raphanus* and *Brassica*. Initially, it was primarily used in cytogenetic studies to achieve the introduction of valuable traits (Karpechenko, 1924; McNaughton, 1973). *Raphanobrassica* contains 4MSOB, which is present in high concentrations in *B. oleracea*, and it contains 4-methylsulfinyl-3-butenyl GSL (glucoraphenin, 4MSO3B), which is in the leaves of *Raphanus* plants (Schutze et al., 1999; Niimi et al., 2015). The ratio of 4MSOB to 4MSO3B content in *Raphanobrassica* is lower; therefore, the development of a cultivar with a relatively higher 4MSOB content is desirable.

The radish *GLUCORAPHASATIN SYNTHASE 1* (*GRS1*) gene encodes a 2-oxoglutarate-dependent dioxygenase that presumably desaturates the 4-(methylthio)butyl GSL (glucoerucin, 4MTB) side chain (Figure 1B). A *GRS1* insertional mutants (*grs1*) are known for their high 4MTB accumulation (Kakizaki et al., 2017). Based on these facts, we hypothesized that metabolism would shift towards 4MSOB synthesis in radish lines containing the *grs1* allele as a parent of *Raphanobrassica* (Figure 1C). This would enable the breeding of *Raphanobrassica* cultivars with high 4MSOB content in the leaves. In this study, allodiploid *Raphanobrassica* (2n = 2x = 18, RC) plants were produced from a cross between kale and radish with heterozygous *GRS1*, and the correlation between *GRS1* genotype and 4MSOB content was analyzed.

2 Methods

2.1 Plant material and growth conditions

A previous study reported a Japanese radish (*Raphanus sativus* L.) landrace, 'cv. Nishimachi-Riso', which contains a *GRS1* mutation (Ishida et al., 2015). As the genotype of *GRS1* was not fixed within the 'cv. Nishimachi-Riso' population, the genotype of *GRS1* was determined using DNA markers (see 2.2 Genotyping) and plants

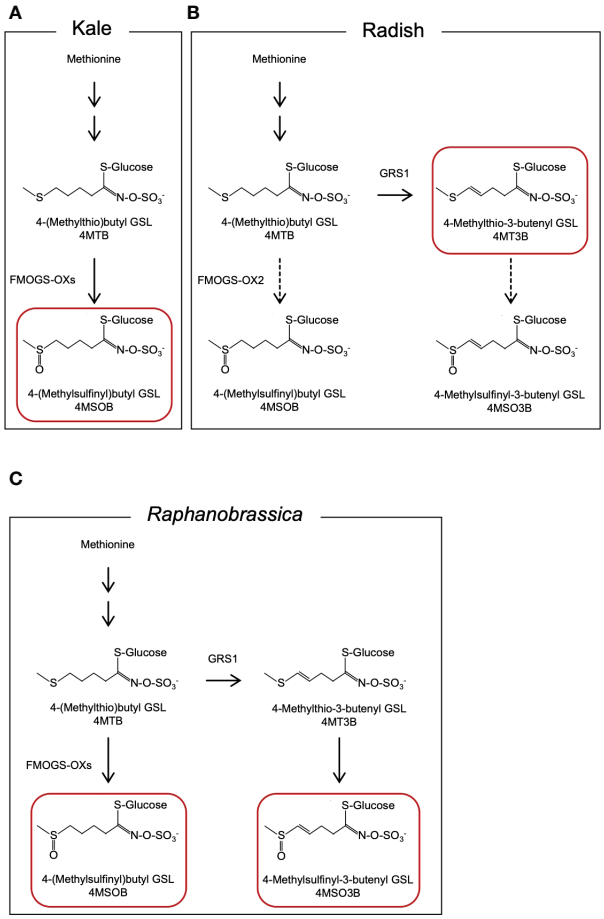


FIGURE 1
Presumed metabolic pathway of aliphatic GSLs in **(A)** kale, **(B)** radish, and **(C)** *Raphanobrassica* leaves. GSL, glucosinolate. The GSLs indicated by red boxes are the major GSLs in each species.

carrying heterozygouse *GRS1* (*GRS1/grs1*) were designated as AKO lines. Kale (*Brassica oleracea* var. *acephala*) inbred line 'KK45-2' is a collard-type kale whose leaves contain 4MSOB used as the paternal parent. All the plants were grown in plastic pots (diameter of 210 mm) in a greenhouse. A Nippi-engei-baido (Nihon Hiryo Co.,Ltd, Gunma, Japan) based soil was used and 1/1000 HYPONeX (HYPONeX JAPAN CORP., LTD., Osaka, Japan) solution was applied once a week. To obtain hybrids, the anthers were removed from radish buds 1–3 d before flowering and were pollinated with kale pollen on the day of flowering. *Raphanobrassica* seeds were sown on September 5, 2020, and genomic DNA was purified from the true leaves two weeks later and used for genotyping. For genomic DNA purification, DNeasy 96 Plant Kit (QIAGEN, Venlo, Netherlands) was used. The plants were then planted in a field on August 31, 2020, with a gap of 35 cm between plants and 60 cm between rows, at the Institute of Vegetable and Floriculture Science (34°46′N, 136°25′E; Tsu, Mie, Japan).

2.2 Genotyping

Previous studies have shown that insertion into the first exon of *GRS1* causes complete functional defects (Kakizaki et al., 2017). To detect the insertion into *GRS1*, primers were designed at the genomic positions shown in Figure 2A, and polymerase chain reaction (PCR) was used to amplify the region between Rs270 (5′-GCAGGAGAGGATGCTTGAAGG-3′) and Rs271 (5′-TGAAACCTTACCCCAAAACG-3′) for the functional type, Rs270 and Rs272 (5′-TCCAGGTTGGGATAGCTTGT-3′) for the defective type. PCR was performed under the following conditions: initial denaturation at 94°C for 1 min, cycling at 94°C for 15 s for heat denaturation, annealing at 60°C for 15 s, and extension at 72°C for 50 s for 32 cycles. The amplified PCR products were separated on a 2% agarose gel, and the functional and defective types were distinguished based on differences in fragment length (Figure 2B). To confirm that the kale genome was inherited by the hybrid, PCR was performed using primers Bo-Fw (5′-CTAGTATGA GGACTCGTTCAGTTACCTCCCTTAGCAGC-3′) and Bo-Rv (5′-GTTTCTTAGAATATGGTGATTGCTGGCTT-3′) to amplify *UDP-sulfoquinovose synthase* (LOC106306866), which is located on chromosome C1.

2.3 GSL analysis

Leaf, root, stem, and flower bud are used for GSL analysis (Figure 3). For leaf analysis, 10 cm from the tip of the 20 cm long leaf was collected, and the central vein was removed. Three leaves per plant were collected and analyzed as one bulk sample. For root analysis, a section was cut 5 cm below the stem/hypocotyl border in the form of a disk (thickness, 0.5–1.0 cm) and collected. For bud analysis, apical flower buds were collected. For stem analysis, a 10 cm section below the apical flower bud of the main stem, which was elongated after bolting, was collected.

After sampling, each part was frozen in liquid nitrogen and dried using a lyophilizer (Labconco, Kansas City, MO, USA). The dried samples were crushed using a multi-bead shocker (Yasui Machinery, Miyazaki, Japan); 0.1 g of ground samples were weighed, mixed with 4.8 mL of 80% methanol and 0.2 mL of 5 mM 2-propenyl GSL (Sigma-Aldrich, St. Louis, MO, USA) as an internal standard and shaken for 30 min at 25°C. After centrifugation at 3,000 rpm for 10 min, the supernatant was collected, GSLs were adsorbed onto DEAE-Sephadex A-25 (Sigma-Aldrich, St. Louis, MO, USA) and desulfonated using arylsulfatase (Type H-1, EC 3.1.6.1, Sigma-Aldrich, St. Louis, MO, USA) at 25°C for 18 h. Desulfo-GSL solutions eluted ion-exchange water were used as samples for high-performance liquid chromatography (HPLC) analysis (LC-20A; Shimadzu Corp., Kyoto, Japan). A reverse-phase column (COSMOSIL 5C18-II, 150 × 4.6 mm; Nacalai Tesque Inc., Kyoto, Japan) was used at 30°C and a flow rate of 1.5 mL/min. The mobile phase comprised 20% acetonitrile, and detection was performed using UV light with a wavelength of 229 nm. The GSL molecular species were estimated based on the retention time of the peaks according to our previous report (Ishida et al., 2015). The individual GSL contents were calculated by the ratios of the individual desulfo-GSL peak areas to the peak areas of an internal standard, 2-propenyl GSL (Sigma-Aldrich, St. Louis, MO, USA), and a response factor (The International Organization for Standardization, 1992). Table 1 shows the list of GSLs analyzed, and Figure 4 shows representative chromatograms of each desulfo-GSL molecular species detected using HPLC.

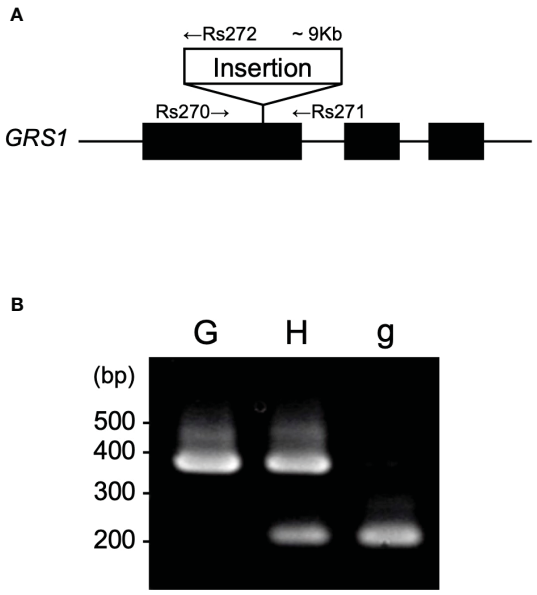


FIGURE 2
(A) Gene structure of *GRS1*. Black boxes represent exons. Arrows represent the position of primers for PCR. **(B)** Representative agarose gel pattern of amplified PCR product. "G", "H", and "g" indicate homozygote for *GRS1*, heterozygote, and homozygote for *grs1*, respectively. *GRS1, GLUCORAPHASATIN SYNTHASE 1*; PCR, polymerase chain reaction.

3 Results

3.1 Intergeneric cross between radish and kale

To generate intergeneric hybrids that harbored the *grs1* allele, we used a radish AKO line that was heterozygous for the *GRS1* gene (*GRS1*/*grs1*) as a seed parent. The kale KK45-2 line, containing 4MSOB in mature leaves, was used as the pollen parent. At the time of crossing, all stamens were removed from radish buds 1–2 d before flowering to avoid self-fertilization, and pollination was performed using kale pollen on the day of flowering. Twenty-one AKO plants were pollinated using KK45-2 pollen, and the pod formation rate varied from 0 to 0.24 among radish plants (Table 2). Similarly, there was a difference in the number of seeds per pod among the AKO plants. Among the 21 combinations, the seeds of four radish plants (AKO103, AKO108, AKO110, and AKO118) that yielded a large number of seeds were sown in petri dishes, and their germination rates and genotypes were investigated (Table 3). The germination percentage ranged from 71.4–97.1%, and 47–87 plants were obtained from each combination. To confirm that the obtained plants were hybrids, PCR was performed using primers specific for kale *UDP-sulfoquinovose synthase* (LOC106306866). This confirmed the inheritance of the kale genome. No amplification of the kale genome was detected in only four plants obtained from AKO110, and these plants were inbred radish plants. Genotyping using the *GRS1* marker (Figure 2) resulted in a 1:1 match for *GRS1* segregation in all combinations. No plants heterozygous for the *GRS1* marker were identified. Based on these results, we obtained hybrids of radish and kale, and *GRS1* segregation followed theoretical values. These data allowed us to evaluate the relationship between the *GRS1* genotype and 4MSOB quantity in the *Raphanobrassica* population.

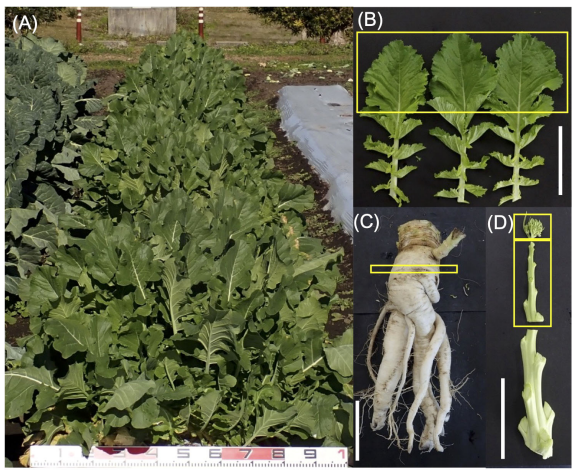


FIGURE 3
(A) Typical phenotypic characteristics of *Raphanobrassica* **(A)**. Plant parts used for glucosinolate analysis: **(B)** leaf, **(C)** root, **(D)** flower bud, and stem. Yellow boxes show a part used for high-performance liquid chromatography analysis. White bars indicate 10 cm.

3.2 Suppression of *GRS1* function increased 4MSOB content in the intergeneric hybrid

The obtained hybrids were planted in a field on September 5, 2020, and the GSL composition in the true leaves (20 cm length) was analyzed 72 d after planting. Growth was vigorous, and leaf shape was similar to that of radish (Figures 3A, B). No plants with pollen fertility were observed for any of the cross combinations. The root shape was snarled and the main root branched into several branches (Figure 3C). The timing of bolting was earlier than that of the parents, and the shape of the flower buds was similar to that of kale (Figure 3D). The 4MSOB content in the true leaves of intergeneric hybrids and parent plants is shown in Figure 5. Notably, in all cross combinations, the *grs1*-type had a 4MSOB content approximately twice as high as that of functional *GRS1* hybrids (Table 4). Furthermore, these *grs1*-type plants had higher 4MSOB concentrations than those in kale KK45-2 and radish AKO103 (*grs1*/*grs1*) plants. There were significant differences in the concentration of 4MSOB between *GRS1* genotypes, but even within the same *GRS1* genotype the concentration of 4MSOB varied widely (Figure 5). The average of 4MSOB in four lines in *GRS1*-type and *grs1*-type were 15.5 ± 0.5 and 34.1 ± 1.0 μmol·g^{-1} dry weight in leaves, respectively. The highest 4MSOB concentration was 69.4 μmol·g^{-1} dry weight. In *grs1*-type *Raphanobrassica*, 4MSO3B was almost undetectable and the accumulation of 4MTB was detected (Figure 6). The contents of indolic GSLs, such as Indol-3-ylmethyl GSL (glucobrassicin, I3M), were not affected by the *GRS1* mutation.

TABLE 1 Information of glucosinolates in the present study.

Peak number[a]	Retention time (min)	Chemical name	Trivial name	Abbreviation	Compound groups	Radish RR genome	Kale CC genome	Raphanobrassica RC genome
1	3.7	3-(Methylthio) propyl	Glucoiberin	3MSOP	Aliphatic		✓[c]	✓
2[b]	5.9	2-Propenyl	Sinigrin	2-Propenyl	Aliphatic	I.C.	I.C.	I.C.
3	6.2	4-(Methylsulfinyl) butyl	Glucoraphanin	4MSOB	Aliphatic		✓	✓
4	6.6	4-Methylsulfinyl-3-butenyl	Glucoraphenin	4MSO3B	Aliphatic	✓		✓
5	10.8	4-Methoxyindol-3-ylmethyl	4-Hydroxyglucobrassicin	4OH-I3M	Indolic		✓	
6	14.4	4-(Methylthio) butyl	Glucoerucin	4MTB	Aliphatic			✓
7	15.2	4-Methylthio-3-butenyl	Glucoraphasatin	4MT3B	Aliphatic	✓		✓
8	16.3	Indol-3-ylmethyl	Glucobrassicin	I3M	Indolic	✓	✓	✓
9	19.1	4-Methoxyindol-3-ylmethyl	4-Methoxyglucobrassicin	4MO-I3M	Indolic	✓	✓	✓

[a]Peak number is same as in Figure 1. [b]2-Propenyl GSL is used for internal control (I.C.). [c]✓ indicates GSL detected.

3.3 GSL content in roots, flower buds, and stems was affected by GRS1 function

The *grs1* mutation increased 4MSOB concentration by approximately 2-fold in the leaves of the hybrids. Next, we analyzed the GSL profiles in various edible plant parts such as roots, flower buds, and stems (Table 5). The 4MSOB content was highest in the flower buds, followed by stems and roots. The *grs1* mutation increased the concentration of 4MSOB by more than 2-fold in flower buds and stems, similar to that in leaves. The major GSLs in the roots were 4MT3B and 4MTB, whereas those in the flower buds and stems were 4MSO3B and 4MSOB (Figure 7). 2-Hydroxy-3-butenyl GSL (progoitrin, 2H3B) and I3M, which are rarely detected in the roots, were detected in flower buds and stems. Regarding the total GSL, the *GRS1*-type plants had a higher content than that in the *grs1*-type, similar to that in the leaves. The total GSL content was highest in flower buds for both *GRS1* and *grs1*-types, followed by roots and stems.

4 Discussion

4.1 An intergeneric hybrid between radish and kale

Raphanobrassica, an intergeneric hybrid between the genera *Raphanus* and *Brassica*, contains large amounts of 4MSOB and 4MSO3B (Schutze et al., 1999). Lyophilized powder derived from *Raphanobrassica* inhibits *Helicobacter pylori*-induced gastritis in Mongolian gerbils (Yamada et al., 2014). Although

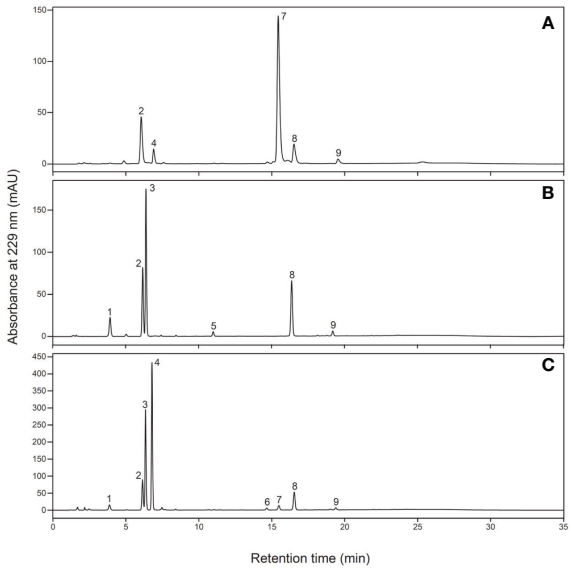


FIGURE 4
Typical chromatograms of desulfo-glucosinolates in **(A)** radish, **(B)** kale, and **(C)** *Raphanobrassica*. Peak 1, 3-(methylthio)propyl; peak 2, 2-propenyl (internal standard); peak 3, 4-(methylsulfinyl)butyl; peak 4, 4-methylsulfinyl-3-butenyl; peak 5, 4-methoxyindol-3-ylmethyl; peak 6, 4-(methylthio)butyl; peak 7, 4-methylthio-3-butenyl; peak 8, indol-3-ylmethyl; peak 9, 4-methoxyindol-3-ylmethyl.

10.3389/fpls.2023.1132302

TABLE 2 Comparative results of intergeneric cross between radish x kale.

♀:Radish line	♂:Kale line	Number of pollinations	Number of sillique developed	Number of seeds obtained	pod/pollinated flower	Seed/pollinated flower
AKO101	KK45-2	422	1	1	0.00	0.00
AKO102	KK45-2	1142	9	11	0.01	0.01
AKO103	KK45-2	595	143	114	0.24	0.19
AKO104	KK45-2	304	2	3	0.01	0.01
AKO105	KK45-2	525	5	5	0.01	0.01
AKO106	KK45-2	517	0	0	0.00	0.00
AKO107	KK45-2	331	0	0	0.00	0.00
AKO108	KK45-2	503	72	105	0.14	0.21
AKO109	KK45-2	694	10	15	0.01	0.02
AKO110	KK45-2	845	57	129	0.07	0.15
AKO111	KK45-2	520	1	1	0.00	0.00
AKO112	KK45-2	898	9	8	0.01	0.01
AKO113	KK45-2	352	0	0	0.00	0.00
AKO114	KK45-2	685	13	22	0.02	0.03
AKO115	KK45-2	311	0	0	0.00	0.00
AKO116	KK45-2	212	0	0	0.00	0.00
AKO117	KK45-2	594	1	1	0.00	0.00
AKO118	KK45-2	515	109	175	0.21	0.34
AKO119	KK45-2	310	0	0	0.00	0.00
AKO120	KK45-2	409	4	4	0.01	0.01
AKO121	KK45-2	19	0	0	0.00	0.00

Raphanobrassica has useful properties, the hybridization rate between *Raphanus* and *Brassica* is considerably low. In 1973, McNaughton reported that the number of seeds per pollinated flower between *R. sativus* and *B. oleracea* is between 0.38 and 0.4 (McNaughton, 1973). This value is comparable to that of the most efficient combination in the crossing experiment of the present study (AKO118 × KK45-2, 0.34). Further, the formation rates of hybrids differ between varieties (Kakizaki, 1925). This phenomenon is known as the "hybridization barrier" and is divided into two types: pre-zygotic and post-zygotic barriers. The pre-zygotic barrier is caused mainly by defects in fertilization, such as interspecific incompatibility and defects in pollen tube guidance (Dresselhaus and Marton, 2009). The post-zygotic barrier includes hybrid embryo breakdown and hybrid sterility. Candidate genes or quantitative trait loci responsible for these barriers have been cloned (Udagawa et al., 2010; Tonosaki et al., 2013). In the present study, the hybridization rates were markedly different between AKO plants originating from the same radish 'cv. Nishimachi-Riso' (Table 2). It is not known whether the barriers observed in the present study are pre- or post-zygotic, but differences in their degree within the same species may provide good material for genetic analysis.

TABLE 3 Germination rate and segregation ratio of *GRS1* in allodiploid *Raphanobrassica*.

Cross combination	Number of seeds	Germination percentage (%)	Plants obtained	Absence of kale marker[a]	Number of hybrid	Genotype with *GRS1* marker			
						GRS1	*grs1*	*GRS1/grs1*	χ^2 (1:1)[b]
AKO103 x KK45-2	88	86.4	47	0	47	18	29	0	2.574
AKO108 x KK45-2	70	97.1	55	0	55	29	26	0	0.164
AKO110 x KK45-2	112	71.4	72	4	68	31	37	0	0.529
AKO118 x KK45-2	113	85.0	87	0	87	46	41	0	0.287

a) UDP-sulfoquinovose synthase (LOC106306866), C01.
b) chi-squared = 3.841, *df* = 1, p = 0.05.

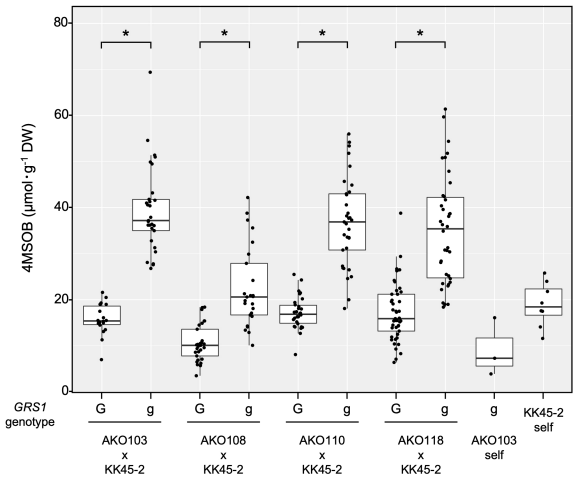


FIGURE 5
Distribution of leaf 4SMOB content in intergeneric hybrids and parent lines. The genotype of *GRS1* is indicated by "G" for functional and "g" for null allele. Asterisks above the bars indicate significant differences ($p < 0.01$) between the genotypes of *GRS1* as determined *via* the Welch's t-test. *GRS1, GLUCORAPHASATIN SYNTHASE 1.*

4.2 GSLs in *Raphanobrassica*

The hybridization between a radish and Chinese cabbage produces a cultivar called 'Baemoochae' (Lee et al., 2011). Baemoochae contains GSL molecular species present in both radish and Chinese cabbage. However, the total GSL content is similar to that of both parents (Nugroho et al., 2020). In contrast, the total GSL content of *Raphanobrassica* produced in this study was higher than that of either parent, suggesting that the entire GSL synthesis pathway was activated. It is reported that the GSL synthesis gene expression is elevated in resynthesized *Brassica* allotetraploids compared with that in their diploids (Zhang et al., 2015). Further studies are needed to determine whether the increased 4MSOB in *Raphanobrassica* developed in this study is solely due to mutations in *GRS1* or is also influenced by the increased expression of other biosynthetic enzyme genes. GSLs are hydrolyzed by myrosinases to not only ITCs but other products (thiocyanates, nitriles, or epithionitriles) depending on pH and the presence of epithiospecifier protein (Bones and Rossiter, 1996). Distribution of degradated products vary widely between plant species. *B. oleracea* has a higher proportion of nitriles and epithionitriles than ITCs, whereas *R. sativus* has a higher proportion of ITCs (Cole, 1976). Therefore, analysis of the abundance ratios of degradation products in *Raphanobrassica*, which has both genomes, is of great importance for its use as a functional vegetable.

4.3 Effects of *grs1* mutation on the GSL synthesis pathway

Hybrids containing *grs1* showed almost no 4MT3B or 4MSO3B content in their leaves (Table 4 and Figure 6). This result supports the hypothesis that the introduction of *grs1* into *Raphanobrassica* enhances 4MTB utilization in 4MSOB synthesis and increases the 4MSOB content. However, in all cross combinations, the total content of each GSL in the population possessing *grs1*-type tended to be lower than that in *GRS1*-type plants (Figure 6). Therefore, increased 4MSOB content in plants lacking GRS1 function may cause feedback inhibition of FMO. For example, YUCCA, an enzyme in the auxin synthesis pathway, also belongs to the monooxygenase family, similar to FMO. *YUCCA* transcription levels are negatively regulated by the synthetic product auxin (Suzuki et al., 2015). If FMO activity and transcription levels are negatively regulated by 4MSOB, understanding the underlying mechanisms and applying them in breeding could lead to higher 4MSOB content.

TABLE 4 Relationship between *GRS1* genotype and glucosinolate content.

Cross combination	*GRS1* genotype	Number of plants	μmol·g^{-1} dry weight		4MSOB ratio[a]
			4MSOB	4MSO3B	
AKO103 x KK45-2	*GRS1*	18	15.8 ± 0.8^{bc}	30.6 ± 1.1^{b}	2.51
	grs1	29	39.6 ± 1.7^{a}	0.04 ± 0.3^{d}	
AKO108 x KK45-2	*GRS1*	29	10.6 ± 0.7^{c}	22.2 ± 1.0^{c}	2.13
	grs1	25	22.6 ± 1.7^{b}	0.2 ± 0.1^{d}	
AKO110 x KK45-2	*GRS1*	30	17.2 ± 0.6^{b}	35.1 ± 1.1^{a}	2.14
	grs1	33	36.8 ± 1.7^{a}	0.3 ± 0.1^{d}	
AKO118 x KK45-2	*GRS1*	45	17.4 ± 1.0^{b}	26.9 ± 1.2^{b}	2.02
	grs1	38	35.1 ± 1.9^{a}	0.2 ± 0.05^{d}	
AKO103 S1	*grs1*	3	9.1 ± 3.0	not detected	
KK45-2 S1	–	8	19.0 ± 1.6	not detected	

Glucosinolate content is indicated as averages ± SE. Values given are μmol g^{-1} dry weight.
Same letters indicate no significant difference (Tukey-Kramer HSD test, $P < 0.05$).
[a] 4MSOB ratio of *GRS1* type to *grs1* type as 1 for hybrids resulting from the same cross combination.

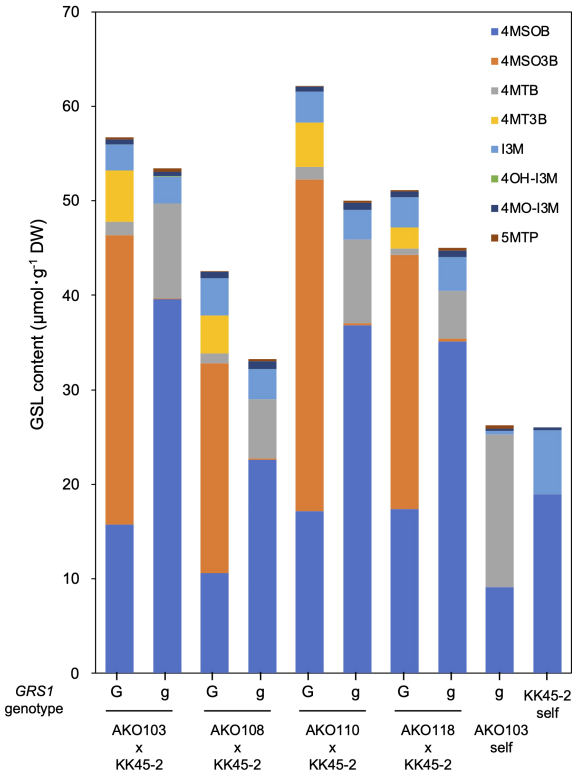


FIGURE 6
GSL composition and content in *Raphanobrassica* leaf. The genotype of *GRS1* is indicated by "G" for functional and "g" for null allele. *GRS1, GLUCORAPHASATIN SYNTHASE 1*; GSL, glucosinolate. 4MSOB, 4-(methylsulfinyl)butyl GSL; 4MSO3B, 4-methylsulfinyl-3-butenyl GSL; 4MTB, 4-(methylthio)butyl GSL; 4MT3B, 4-methylthio-3-butenyl GSL; I3M, indol-3-ylmethyl GSL; 4OH-I3M, 4-methoxyindol-3-ylmethyl GSL, 4MO-I3M, 4-methoxyindol-3-ylmethyl GSL; 5MTP, 5-(methylthio)pentyl GSL.

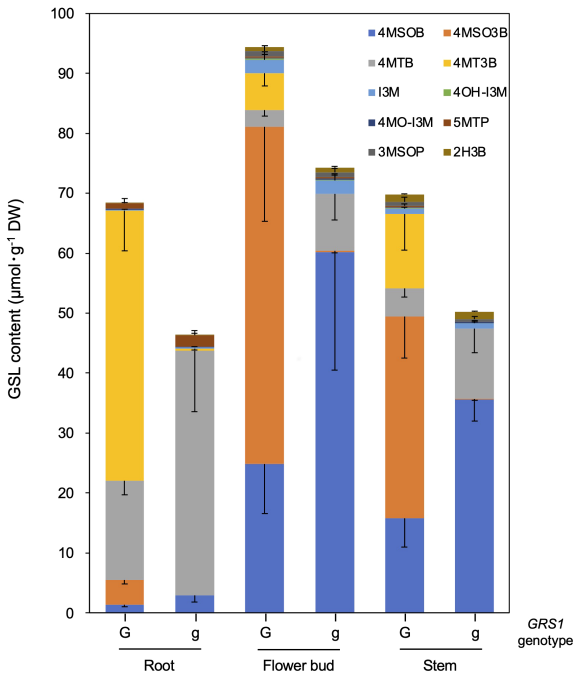


FIGURE 7
GSL composition and content in root, flower bud, and stem in *Raphanobrassica*. The genotype of *GRS1* is indicated by "G" for functional and "g" for null allele. Error bars represent standard deviation. *GRS1, GLUCORAPHASATIN SYNTHASE 1*; GSL, glucosinolate. 4MSOB, 4-(methylsulfinyl)butyl GSL; 4MSO3B, 4-methylsulfinyl-3-butenyl GSL; 4MTB, 4-(methylthio)butyl GSL; 4MT3B, 4-methylthio-3-butenyl GSL; I3M, indol-3-ylmethyl GSL; 4OH-I3M, 4-methoxyindol-3-ylmethyl GSL, 4MO-I3M, 4-methoxyindol-3-ylmethyl GSL; 5MTP, 5-(methylthio)pentyl GSL; 3MSOP, 3-(methythio)propyl GSL; 2H3B, 2-hydroxy-3-butenyl GSL.

4.4 Potential and challenges of using *Raphanobrassica* as a high-4MSOB containing vegetable

4MSOB is present in the genera *Brassica*, *Eruca, and Raphanus* (Ciska et al., 2000; Ishida et al., 2014). A study in 1992 reported that sulforaphane (ITC derived from 4MSOB) in broccoli functions as a major inducer of anti-carcinogenic defense enzymes. Consequently, the recognition of broccoli as a representative vegetable containing 4MSOB has increased, which has prompted research in various fields. To develop broccoli varieties with high 4MSOB content, *MYB28* of the wild species *B. villosa* (*BvMYB28*) has been introduced into cultivated species, and the F_1 hybrid Beneforté® has been cultivated (Traka et al., 2013). In *Arabidopsis*, *AtMYB28* positively regulates aliphatic GSL biosynthesis (Sonderby et al.,

TABLE 5 Glucosinolate content in different organ of *Raphanobrassica*.

Cross combination	Organ	*GRS1* genotype	Number of plants	μmol·g^{-1} dry weight		4MSOB ratio[a]
				4MSOB	4MSO3B	
AKO103 x KK45-2	Root	*GRS1*	10	1.4 ± 0.1^{d}	4.1 ± 0.2^{c}	2.07
		grs1	10	2.9 ± 0.4^{d}	not detected	
AKO110 x KK45-2	Flower bud	*GRS1*	10	24.8 ± 2.6^{bc}	56.2 ± 5.0^{a}	2.42
		grs1	10	60.1 ± 6.2^{a}	0.2 ± 0.1^{c}	
AKO110 x KK45-2	Stem	*GRS1*	10	15.8 ± 1.5^{c}	33.6 ± 2.2^{b}	2.25
		grs1	10	35.6 ± 1.1^{b}	0.1 ± 0.1^{c}	

Glucosinolate content is indicated as averages ± SE. Values given are μmol g^{-1} dry weight.
Same letters indicate no significant difference (Tukey-Kramer HSD test, P < 0.05).
[a] 4MSOB ratio of *GRS1* type to *grs1* type as 1 for hybrids resulting from the same cross combination.

2007). The expression of several aliphatic GSL genes is elevated in broccoli, which is homozygous for *BvMYB28*. The 4MSOB concentration in the homozygous *BvMYB28* inbred broccoli line is 20 μmol·g^{-1} dry weight in floret (Neequaye et al., 2022). In contrast, the average value of 4MSOB in the leaves of *Raphanobrassica* with *grs1* grown in this study was 34.1 μmol·g^{-1} dry weight (Table 4). Although simple comparisons cannot be made because the analyzed sites contained the highest contents in adult plants and the content varies with the growing region and cultivation method, the variety with defective GRS1 in *Raphanobrassica* may be used in various cuisines as a new leafy vegetable that contains as much 4MSOB in broccoli Beneforté®. Compared with kale, the taste of *Raphanobrassica* leaves is softer and less gruel-like, making it suitable for a variety of dishes such as stir-fry and salads. The low hybridization affinity between radish and kale and the inability to ensure a commercial level of seed production represent issues that need to be addressed to popularize the variety in the future. However, chromosome doubling may be used to breed seed-fertile allotetraploid *Raphanobrassica* (Niimi et al., 2015).

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

TK and HC supervised and conceived the project. RE and TK wrote the manuscript. EI, MK, and TO supported the experiments and revised the manuscript. TK and RE cultivated the plants and evaluated the GSLs. All authors contributed to the article and approved the submitted version.

Funding

Research funding for this study was provided in part by KAGOME CO., LTD.

Acknowledgments

We thank Y. Kawamoto, S. Morimoto, E. Yamamoto, H. Saito, and M. Kitazumi for technical assistance. We would like to thank Editage (www.editage.com) for English language editing.

Conflict of interest

KAGOME CO., LTD. Nagoya, Japan supported this work and provided support in the form of salaries to RE and HC.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1132302/full#supplementary-material

References

Bones, A. M., and Rossiter, J. T. (1996). The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiol. Plant.* 97, 194–208. doi: 10.1111/j.1399-3054.1996.tb00497.x

Cartea, M. E., and Velasco, P. (2007). Glucosinolates in brassica foods: bioavailability in food and significance for human health. *Phytochem. Rev.* 7, 213–229. doi: 10.1007/s11101-007-9072-2

Chen, Q. M., and Maltagliati, A. J. (2018). Nrf2 at the heart of oxidative stress and cardiac protection. *Physiol. Genomics* 50, 77–97. doi: 10.1152/physiolgenomics.00041.2017

Ciska, E., Martyniak-Przybyszewska, B., and Kozlowska, H. (2000). Content of glucosinolates in cruciferous vegetables grown at the same site for two years under different climatic conditions. *J. Agric. Food Chem.* 48, 2862–2867. doi: 10.1021/jf981373a

Cole, R. A. (1976). Isothiocyanates, nitriles and thiocyanates as products of autolysis of glucosinolates in *Cruciferae*. *Phytochemistry* 15, 759–762. doi: 10.1016/S0031-9422(00)94437-6

Dresselhaus, T., and Marton, M. L. (2009). Micropylar pollen tube guidance and burst: adapted from defense mechanisms? *Curr. Opin. Plant Biol.* 12, 773–780. doi: 10.1016/j.pbi.2009.09.015

Egner, P. A., Chen, J. G., Zarth, A. T., Ng, D. K., Wang, J. B., Kensler, K. H., et al. (2014). Rapid and sustainable detoxication of airborne pollutants by broccoli sprout beverage: results of a randomized clinical trial in China. *Cancer Prev. Res. (Phila)* 7, 813–823. doi: 10.1158/1940-6207.CAPR-14-0103

Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51. doi: 10.1016/S0031-9422(00)00316-2

Farnham, M. W., Wilson, P. E., Stephenson, K. K., and Fahey, J. W. (2004). Genetic and environmental effects on glucosinolate content and chemoprotective potency of broccoli. *Plant Breed.* 123, 60–65. doi: 10.1046/j.0179-9541.2003.00912.x

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Ishida, M., Hara, M., Fukino, N., Kakizaki, T., and Morimitsu, Y. (2014). Glucosinolate metabolism, functionality and breeding for the improvement of brassicaceae vegetables. *Breed Sci.* 64, 48–59. doi: 10.1270/jsbbs.64.48

Ishida, M., Kakizaki, T., Morimitsu, Y., Ohara, T., Hatakeyama, K., Yoshiaki, H., et al. (2015). Novel glucosinolate composition lacking 4-methylthio-3-butenyl glucosinolate in Japanese white radish *(Raphanus sativus* l.). *Theor. Appl. Genet.* 128, 2037–2046. doi: 10.1007/s00122-015-2564-3

Kakizaki, Y. (1925). A preliminary report of crossing experiments with cruciferous plants, with special reference to sexual compatibility and matroclinous hybrids. *Japanese J. Genet.* 3, 49–82. doi: 10.1266/jjg.3.49

Kakizaki, T., Kitashiba, H., Zou, Z., Li, F., Fukino, N., Ohara, T., et al. (2017). A 2-Oxoglutarate-Dependent dioxygenase mediates the biosynthesis of glucoraphasatin in radish. *Plant Physiol.* 173, 1583–1593. doi: 10.1104/pp.16.01814

Karpechenko, G. D. (1924). Hybrids of female *Raphanus sativus* l x male *Brassica oleracea* l. *J. Genet.* 14, 375–396. doi: 10.1007/BF02983104

Kensler, T. W., Chen, J. G., Egner, P. A., Fahey, J. W., Jacobson, L. P., Stephenson, K. K., et al. (2005). Effects of glucosinolate-rich broccoli sprouts on urinary levels of aflatoxin-DNA adducts and phenanthrene tetraols in a randomized clinical trial in he zuo township, qidong, people's republic of China. *Cancer Epidemiol. Biomarkers Prev.* 14, 2605–2613. doi: 10.1158/1055-9965.EPI-05-0368

Kensler, T. W., Egner, P. A., Agyeman, A. S., Visvanathan, K., Groopman, J. D., Chen, J. G., et al. (2013). Keap1-nrf2 signaling: a target for cancer prevention by sulforaphane. *Top. Curr. Chem.* 329, 163–177. doi: 10.1007/128_2012_339

Kikuchi, M., Ushida, Y., Shiozawa, H., Umeda, R., Tsuruya, K., Aoki, Y., et al. (2015). Sulforaphane-rich broccoli sprout extract improves hepatic abnormalities in male subjects. *World J. Gastroenterol.* 21, 12457–12467. doi: 10.3748/wjg.v21.i43.12457

Lee, S. S., Lee, S. A., Yang, J., and Kim, J. (2011). Developing stable progenies of *xBrassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. *Theor. Appl. Genet.* 122, 885–891. doi: 10.1007/s00122-010-1494-3

McNaughton, I. H. (1973). Synthesis and sterility of raphanobrassica. *Euphytica* 22, 70–88. doi: 10.1007/BF00021558

Neequaye, M., Steuernagel, B., Saha, S., Trick, M., Troncoso-Rey, P., Van Den Bosch, F., et al. (2022). Characterisation of the introgression of *Brassica villosa* genome into broccoli to enhance methionine-derived glucosinolates and associated health benefits. *Front. Plant Sci.* 13, 855707. doi: 10.3389/fpls.2022.855707

Niimi, H., Watanabe, M., Serizawa, H., Koba, T., Nakamura, I., and Mii, M. (2015). Amiprophosmethyl-induced efficient *in vitro* production of polyploids in raphanobrassica with the aid of aminoethoxyvinylglycine (AVG) in the culture medium. *Breed Sci.* 65, 396–402. doi: 10.1270/jsbbs.65.396

Nouchi, R., Hu, Q. Q., Saito, T., Kawata, N. Y. D., Nouchi, H., and Kawashima, R. (2021). Brain training and sulforaphane intake interventions separately improve cognitive performance in healthy older adults, whereas a combination of these interventions does not have more beneficial effects: evidence from a randomized controlled trial. *Nutrients* 13, 352. doi: 10.3390/nu13020352

Nugroho, A. B. D., Han, N., Pervitasari, A. N., Kim, D. H., and Kim, J. (2020). Differential expression of major genes involved in the biosynthesis of aliphatic glucosinolates in intergeneric baemoochae *(Brassicaceae)* and its parents during development. *Plant Mol. Biol.* 102, 171–184. doi: 10.1007/s11103-019-00939-2

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in *Brassicaceae*. *Plant Mol. Biol.* 42, 93–113. doi: 10.1023/A:1006380021658

Riedl, M. A., Saxon, A., and Diaz-Sanchez, D. (2009). Oral sulforaphane increases phase II antioxidant enzymes in the human upper airway. *Clin. Immunol.* 130, 244–251. doi: 10.1016/j.clim.2008.10.007

Schutze, W., Mandel, F., and Schulz, H. (1999). Identification of glucosinolates in radish *(Raphanus sativus* l.) and cross-breeds of *R. sativus* l. x *Brassica oleracea* l. (Raphanobrassica) by LC-MS. *Nahrung-Food* 43, 245–248. doi: 10.1002/(SICI)1521-3803(19990801)43:4<245::AID-FOOD245>3.0.CO;2-#

Sonderby, I. E., Hansen, B. G., Bjarnholt, N., Ticconi, C., Halkier, B. A., and Kliebenstein, D. J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One* 2, e1322. doi: 10.1371/journal.pone.0001322

Suzuki, M., Yamazaki, C., Mitsui, M., Kakei, Y., Mitani, Y., Nakamura, A., et al. (2015). Transcriptional feedback regulation of *YUCCA* genes in response to auxin levels in *Arabidopsis*. *Plant Cell Rep.* 34, 1343–1352. doi: 10.1007/s00299-015-1791-z

The International Organization for Standardization (1992). Rapeseed-determination of glucosinolate content. *ISO* 9167–1, 1–9.

Tonosaki, K., Michiba, K., Bang, S. W., Kitashiba, H., Kaneko, Y., and Nishio, T. (2013). Genetic analysis of hybrid seed formation ability of *Brassica rapa* in intergeneric crossings with *Raphanus sativus*. *Theor. Appl. Genet.* 126, 837–846. doi: 10.1007/s00122-012-2021-5

Traka, M. H., Saha, S., Huseby, S., Kopriva, S., Walley, P. G., Barker, G. C., et al. (2013). Genetic regulation of glucoraphanin accumulation in Beneforté® broccoli. *New Phytol.* 198, 1085–1095. doi: 10.1111/nph.12232

Udagawa, H., Ishimaru, Y., Li, F., Sato, Y., Kitashiba, H., and Nishio, T. (2010). Genetic analysis of interspecific incompatibility in *Brassica rapa*. *Theor. Appl. Genet.* 121, 689–696. doi: 10.1007/s00122-010-1340-7

Wermter, N. S., Rohn, S., and Hanschen, F. S. (2020). Seasonal variation of glucosinolate hydrolysis products in commercial white and red cabbages (Brassica oleracea var. *capitata*). *Foods* 9, 1682. doi: 10.3390/foods9111682

Yagishita, Y., Fahey, J. W., Dinkova-Kostova, A. T., and Kensler, T. W. (2019). Broccoli or sulforaphane: is it the source or dose that matters? *Molecules* 24, 3593. doi: 10.3390/molecules24193593

Yamada, T., Wei, M., Toyoda, T., Yamano, S., and Wanibuchi, H. (2014). Inhibitory effect of raphanobrassica on helicobacter pylori-induced gastritis in Mongolian gerbils. *Food Chem. Toxicol.* 70, 107–113. doi: 10.1016/j.fct.2014.04.037

Yuanfeng, W., Chengzhi, L., Ligen, Z., Juan, S., Xinjie, S., Yao, Z., et al. (2021). Approaches for enhancing the stability and formation of sulforaphane. *Food Chem.* 345, 128771. doi: 10.1016/j.foodchem.2020.128771

Zhang, D., Pan, Q., Cui, C., Tan, C., Ge, X., Shao, Y., et al. (2015). Genome-specific differential gene expressions in resynthesized *Brassica* allotetraploids from pair-wise crosses of three cultivated diploids revealed by RNA-seq. *Front. Plant Sci.* 6, 957. doi: 10.3389/fpls.2015.00957

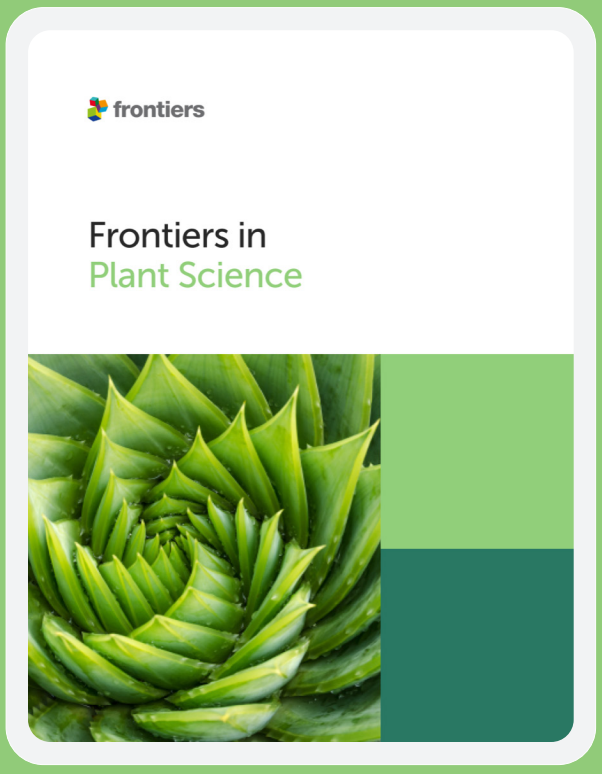
Frontiers in Plant Science

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