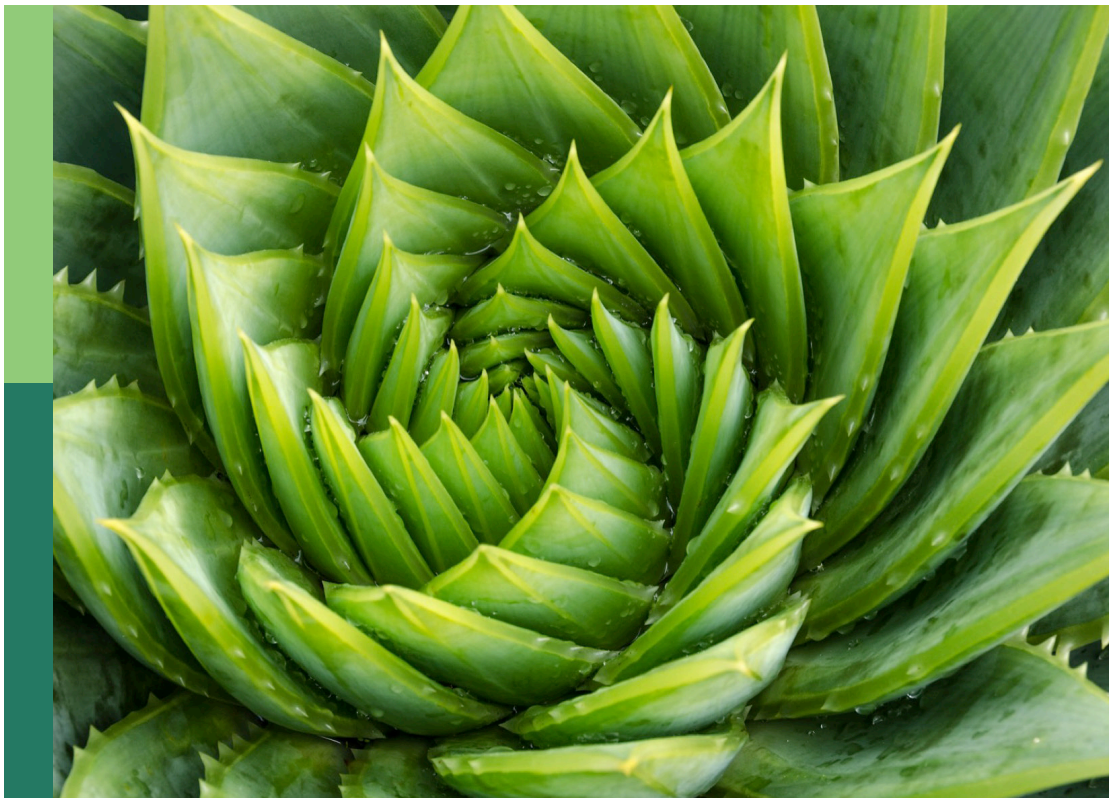
frontiers | Research Topics

Cucurbitaceae: Multi-omics, functional analysis, and molecular breeding

Edited by
Qiusheng Kong, Jie Zhang, Yi Zheng and Yang Bai

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3730-5
DOI 10.3389/978-2-8325-3730-5

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Cucurbitaceae: Multi-omics, functional analysis, and molecular breeding

Topic editors

Qiusheng Kong — Huazhong Agricultural University, China
Jie Zhang — Beijing Academy of Agricultural and Forestry Sciences, China
Yi Zheng — Beijing University of Agriculture, China
Yang Bai — Rensselaer Polytechnic Institute, United States

Citation

Kong, Q., Zhang, J., Zheng, Y., Bai, Y., eds. (2023). *Cucurbitaceae: Multi-omics, functional analysis, and molecular breeding*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3730-5

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 03 October 2023
DOI 10.3389/fpls.2023.1301212

OPEN ACCESS

EDITED AND REVIEWED BY
Nunzio D'Agostino,
University of Naples Federico II, Italy

*CORRESPONDENCE
Qiusheng Kong
qskong@mail.hzau.edu.cn

RECEIVED 24 September 2023
ACCEPTED 26 September 2023
PUBLISHED 03 October 2023

CITATION
Kong Q, Zheng Y, Zhang J and Bai Y (2023) Editorial: Cucurbitaceae: multi-omics, functional analysis, and molecular breeding.
Front. Plant Sci. 14:1301212.
doi: 10.3389/fpls.2023.1301212

Editorial: Cucurbitaceae: multi-omics, functional analysis, and molecular breeding

Qiusheng Kong[1]*, Yi Zheng[2], Jie Zhang[3] and Yang Bai[4]

[1]National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, China, [2]College of Plant Science and Technology, Beijing University of Agriculture, Beijing, China, [3]Beijing Academy of Agricultural and Forestry Sciences, Beijing, China, [4]Rensselaer Polytechnic Institute, Troy, NY, United States

KEYWORDS

Cucurbitaceae, multi-omics, functional analysis, molecular breeding, editorial

Editorial on the Research Topic
Cucurbitaceae: multi-omics, functional analysis, and molecular breeding

As the second largest vegetable group, cucurbits are important economic crops and enjoy great popularity worldwide. After a complex domestication history, there are now a wide variety of cucurbit vegetables, including watermelon, melon, pumpkin, cucumber, bitter melon, bitter gourd, wax gourd, and many other species. In recent years, climate change has exacerbated a series of environmental stresses, such as extreme temperatures, drought, salinity, pests and diseases, which would greatly affect global crop production. Increasing attention has been placed on diverse germplasm resources to explore potential varieties with favorable agronomic traits for cucurbit breeding. Technological advances have enabled a comprehensive knowledge of the physiological and biochemical processes, molecular regulation, and genetic information at the multi-omics level for Cucurbitaceae. To address these problems, this Research Topic focus on the latest findings to describe the complex molecular mechanisms controlling the growth, development, and stress tolerance in Cucurbitaceae plants from a systematic biology perspective. The main goal of this Research Topic is to shed light on the progress made in multi-omics, functional analysis, and molecular breeding for cucurbit crops.

Several loci controlling important traits were genetically mapped for cucurbit crops and the candidate genes were proposed. Nie et al. assembled the parental genomes of a F_2 population and identified 15 quantitative trait loci (QTL) for fruit quality-related traits in watermelon using a high-density genetic map. Sequence variation analysis showed that a gene encoding phytoene synthase was the candidate gene regulating watermelon flesh color changes. Yi et al. found that white flesh trait of watermelon was controlled by a single recessive locus and referred to as *Clwf2*. They mapped the *Clwf2* locus to a 132.3 kb region on chromosome 6 and the gene encoding a tetratricopeptide repeat protein was proposed as the candidate gene. Zhan et al. constructed a recombinant inbred line population containing 164 lines and detected a major QTL controlling watermelon fruit cracking on chromosome 2 using a high-density genetic map. Gong et al. identified five significant

single nucleotide polymorphisms closely associated with watermelon rind thickness on chromosome 2 by genome-wide association study and a MADS family gene was proposed as the candidate gene. Zhu et al. found that leaf yellowing of watermelon was controlled by a single recessive gene and mapped the controlling gene in a region of 2.217 Mb on chromosome 2. Moreover, Liu et al. constructed an interspecific high-density genetic linkage map for *Luffa acutangula* and *Luffa cylindrica* for the first time and identified 40 QTLs related to 25 important agronomic traits. These results provide valuable information for maker development and target gene cloning.

Several genes and pathways related to stress response were characterized for cucurbit crops. Zhang et al. identified 14 members of the *Mildew Locus O* (*MLO*) gene family across melon genome and found *CmMLO5* (*MELO3C012438*) gene played a negative role in regulating powdery mildew-resistance in the susceptible melon line. Cai et al. analyzed the defense responses of wax gourd to wilt disease caused by *Phytophthora melonis* infection and found that phenylpropanoid biosynthesis, plant-pathogen interaction, the mitogen-activated protein kinase signaling pathway and plant hormone signal transduction were the most relevant pathways during the response of wax gourd to *P. melonis* infection. Luo et al. identified seven *DELLA* genes in pumpkin and revealed their potential roles under development and abiotic stresses. These genes and pathways undoubtedly deepen our understanding on the resistance or tolerance of cucurbit crops to biotic and abiotic stresses.

Germplasm innovation was performed for cucurbit crops. Yu et al. used ethyl methanesulfonate to mutate 10,000 seeds to construct a mutant collection for bitter melon and identified 14 lines with mutated phenotypes. Yan et al. identified a cucumber albino *alc* mutant and found it was recessively controlled by a single locus. They further investigated the pathways changed in the albino mutant using transcriptome analysis. These novel germplasms provide essential resources for genetic studies and breeding of cucurbit crops.

Multi-omics approach was used to investigate the regulating mechanism of important traits. Xie et al. identified 31 differentially accumulated flavonoids and 828 differentially expressed genes between the *hfc12* mutant with light-color pericarp and wild-type 'BWT' of wax gourd. Then, they found that the light-color pericarp and higher flavonoid content was controlled by a single gene *BhiPRR6* (*Bhi12M000742*) by using bulked segregant analysis-seq and kompetitive allele specific PCR analysis. Deng et al. explored the molecular fluctuations in pumpkin leaves at different time intervals after foliar Zn treatment using the combination of metabolome and transcriptome, and then they elucidated the molecular mechanisms underlying the foliar Zn application induced increase of 2-acetyl-1-pyrroline conferring the 'taro-like' aroma in pumpkin leaves. Moreover, Yue et al. used the whole-transcriptome analyses to identify key differentially expressed mRNAs, long non-coding RNAs, and microRNA associated with male sterility in watermelon. The multi-omics approach provides us a comprehensive understanding on the regulating mechanism of important traits for cucurbit crops from a systematic biology perspective.

At last, as Guest Editors, we would like to thank all the authors and co-authors for their valuable contributions to this Research Topic, all the reviewers for their important work in evaluating the submitted manuscripts. We expect this Research Topic to contribute to the advancement of genetic improvement for cucurbit crops and believe it serves as a valuable reference to our globally colleagues.

Author contributions

QK: Writing – original draft. YZ: Writing – review & editing. JZ: Writing – review & editing. YB: Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. QK acknowledges support from the Key R&D project of Hubei Province (2021BBA101) and the Fundamental Research Funds for the Central Universities (2662020YLPY024).

Acknowledgments

The authors would like to thank all the contributors of the Research Topic, reviewers, and the Editorial Office of Frontiers in Plant Sciences, whose efforts have led to the success of this Research Topic.

Conflict of interest

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

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fpls.2022.1019787

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Guangwei Zhao,
Zhengzhou Fruit Research Institute, (CAAS), China
Shuping Qu,
Northeast Agricultural University, China
Shaohua Zeng,
South China Botanical Garden, (CAS), China

*CORRESPONDENCE
Cheng Xiong
xiongchenghzau@163.com
Xiaowu Sun
sunxiaowu2007@126.com
Baobin Mi
mibaobin@hunaas.cn

†These authors share first authorship

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

RECEIVED 15 August 2022
ACCEPTED 02 September 2022
PUBLISHED 26 September 2022

CITATION
Xie L, Wang J, Liu F, Zhou H, Chen Y, Pan L, Xiao W, Luo Y, Mi B, Sun X and Xiong C (2022) Integrated analysis of multi-omics and fine-mapping reveals a candidate gene regulating pericarp color and flavonoids accumulation in wax gourd (*Benincasa hispida*). *Front. Plant Sci.* 13:1019787. doi: 10.3389/fpls.2022.1019787

Integrated analysis of multi-omics and fine-mapping reveals a candidate gene regulating pericarp color and flavonoids accumulation in wax gourd (*Benincasa hispida*)

Lingling Xie[1,2†], Jin Wang[1,3†], Feng Liu[1], Huoqiang Zhou[2], Ying Chen[2], Luzhao Pan[1,3], Wei Xiao[2], Yin Luo[2], Baobin Mi[2*], Xiaowu Sun[1*] and Cheng Xiong[1*]

[1]College of Horticulture, Hunan Agricultural University, Changsha, China, [2]Vegetable Research Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [3]College of Horticulture, Nanjing Agricultural University, Nanjing, China

Wax gourd (*Benincasa hispida*), a popular fruit of the Cucurbitaceae (cucurbits) family, contains many nutrients with health benefits and is widely grown in China and other tropical areas. In this study, a wax gourd mutant *hfc12* with light-color pericarp was obtained through ethane methylsulfonate (EMS) mutagenesis. Integrative analysis of the metabolome and transcriptome identified 31 differentially accumulated flavonoids (DAFs; flavonoids or flavonoid glycosides) and 828 differentially expressed genes (DEGs) between the *hfc12* mutant and wild-type 'BWT'. Furthermore, BSA-seq and kompetitive allele specific PCR (KASP) analysis suggested that the light-color pericarp and higher flavonoid content was controlled by a single gene *BhiPRR6* (*Bhi12M000742*), a typical two-component system (TCS) pseudo-response regulator (PRR). Genetic analysis detected only one nonsynonymous mutation (C-T) in the second exon region of the *BhiPRR6*. Weighted correlation network analysis (WGCNA) identified the downstream target genes of *BhiPRR6*, probably regulated by light and were intermediated in the regulatory enzyme reaction of flavonoid biosynthetic pathway. Thus, these results speculated that the transcription factor *BhiPRR6*, interacting with multiple genes, regulates the absorption of light signals and thereby changes the pericarp color and synthesis of flavonoids in wax gourd.

KEYWORDS

wax gourd, flavonoid biosynthesis, fine-mapping, *BhiPRR6*, pericarp color

Introduction

Wax gourd [*Benincasa hispida* (Thunb.) Cogn., 2n = 2x = 24] is an important vegetable of the Cucurbitaceae family widely grown in the tropical and subtropical regions of the world (Xie et al., 2019). The plant bears large fruit with a dark-green pericarp and has high nutritional values, such as proteins, carbohydrates, vitamins and minerals (Sreenivas et al., 2011). As an important commodity value feature, peel color is a highly variable trait controlled by a relatively complex genetic mechanism, and mainly depends on the content and composition of flavonoids (Li et al., 2013).

As one of the major components of wax gourd, the flavonoids have attracted breeders and consumers for their ability to improve metabolic disorders and antiangiogenic and anticancer effects (Zoratti et al., 2014). Besides increasing the nutritional value of the fruit, flavonoids act as indicators of fruit quality also in wax gourd (Imran et al., 2019). Flavonoids are polyphenolic secondary metabolites of plants with different functions and include flavonols, flavones, flavanones, flavan-3-ols, isoflavones, and anthocyanins. They are generally found in plant-based foods, such as fruits, vegetables, beans, and tea, in the bound (flavonoid glycosides) or free form (flavonoid aglycones) (Azuma et al., 2012). Several studies have shown flavonoids have high nutritional functions in the reaction of the organism such as the antioxidant, anti-inflammatory, and antitumor properties (Jaakola and Hohtola, 2010). Notably, flavonoids play a central role in fruit quality and economic value; they influence the color, aroma, astringency, and antioxidant properties of fruits. For example, flavonols provide photoprotective effects, flavan-3-ols (precursors of proanthocyanidins) influence fruit flavor (Sudheeran et al., 2021), and the degree of co-pigmentation of anthocyanins and flavanones affected the fruit color (Baranac et al., 1997). Therefore, it is necessary to understand the factors that regulate the biosynthesis of wax gourd flavonoids, which may help develop and produce improved wax gourd varieties.

Flavonoids are synthesized through the phenylpropanoid and flavonoid pathways. In recent years, the biosynthetic pathway of flavonoids and the structural genes encoding the flavonoid biosynthetic enzymes have been extensively studied (Martens et al., 2010). In the early stages of the flavonoids biosynthetic pathway, the phenylalanine ammonia-lyase (PAL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), and coumadin CoA ligase (4CL) were flavonoid biosynthetic enzymes to produced p-coumaroyl-CoA (precursor substance) (Winkel-Shirley, 2002). Naringenin, catalyzed by chalcone synthase (CHS) and chalcone isomerase (CHI), is the core intermediate that leads to different flavonoid subclasses from each branch in the pathway (Jaakola, 2013). The biosynthesis of anthocyanins is a major branch of the flavonoid pathway, regulated by dihydroflavonol 4-reductase (DFR), UDP-glucose: flavonoid 3-glucosyltransferase (UFGT), and anthocyanin synthase (ANS) (Almeida et al., 2007). Anthocyanin reductase (ANR) and leucoanthocyanidin reductase (LAR) play central roles in proanthocyanidin biosynthesis (Shi et al., 2020). Meanwhile, flavonol is synthesized from dihydroflavonols by flavonol synthase (FLS). Flavones use flavanones as substrate and are formed by flavone synthase (FS) (Tomás-Barberán et al., 2001).

In addition to the structural genes encoding the enzymes of the flavonoid pathway, transcriptional factors regulate the expression of genes encoding these enzymes in each synthesis step. Many transcription factors have been found to regulate the flavonoid pathway (Yan et al., 2021). In general, the transcription factor R2R3-MYB interacts with bHLH (MYC-like basic helix-loop-helix) protein to form a complex with WD40 protein (MBW complex), which regulates the coordinated transcription of structural genes of the flavonoid biosynthetic pathway (Hichri et al., 2011). In apple (*Malus* × *domestica* Borkh.), *MdMYB73* regulates the accumulation and vacuolar acidification of anthocyanins by directly activating the vacuolar transporters *(MdVHA-B1*, *MdVHA-E*, *MdVHP1*, and *MdtDT*) (Hu et al., 2017). In tomato (*Solanum lycopersicum*), *SlMYB14* was involved in the regulation of flavonoid biosynthesis, and played a role in maintaining plant active oxygen homeostasis (Li et al., 2021). In mango (*Mangifera indica* L.), *MiMYB1* acts as a critical regulator of anthocyanin biosynthesis and regulates the light-dependent red coloration (Kanzaki et al., 2020). *PpMYB17*, a bHLH or WD40 cofactor in the MBW complex, activates the structural genes *PpCHS*, *PpCHI*, *PpF3H*, and *PpFLS* involved in flavonoid biosynthetic pathway, and was positively regulates flavonoid biosynthesis in pear (*Pyrus* spp.) fruits (Premathilake et al., 2020). Studies have also reported other transcription factors that regulate flavonoid synthesis. For example, *SQUAMOSA MADS-box* gene, a critical regulatory factor in bilberry (*Vaccinium myrtillus*) fruits, is involved in anthocyanin biosynthesis (Jaakola et al., 2019). In muskmelon (*Cucumis melo*), a *CmKFB* gene negatively regulates flavonoid accumulation (Feder et al., 2015). In recent years, metabolome and transcriptome analysis have shown that *R2R3-MYB*, *bHLH51* and *WRKY23* were candidate key transcription factors that regulate the biosynthesis of flavonoids in cucumber peel (Chen et al., 2021). In wax gourd, a single gene locus on chromosome five controlled peel color (Jiang et al., 2015). However, the transcription factors that regulate the flavonoids synthesis of wax gourd have not been reported.

Ethyl methanesulfonate (EMS), as a chemical mutagen, has been widely used in plant mutagenesis breeding and functional genome research due to its high mutagenesis efficiency, easy operation and wide mutagenesis range. Therefore, EMS mutagenesis can not only obtain deletion or gain-of-function mutants, but also help to understand the role of specific amino

acid residues in protein function (Yuan et al., 2014). BSA (Bulked Segregant Analysis), also known as mixed grouping analysis or cluster segregation analysis, is very suitable for analyzing mutants induced by EMS mutagenesis. Two materials with significant differences are used as parents, and a segregated population is constructed by hybridization. About 30 plants with extreme phenotypes were selected to construct two mixed pools. The difference region obtained by sequencing and analysis of the two mixed pools was the candidate region, and the target gene may exist in this region (Kurlovs et al., 2019). In this study, EMS mutagenesis generated a wax gourd mutant *hfc12* with light peel color and high flavonoid content. The transcriptome sequencing (RNA-seq) and the metabolome (liquid chromatography with tandem mass spectrometry, LC–MS/MS) of the *hfc12* mutant and WT 'BWT' were analyzed to understand the regulation of flavonoid biosynthesis. A single gene *BhiPRR6* (pseudo-response regulator, *Bhi12G000742*) located on the chromosome 12 was fine-mapped by genetic BSA-seq analysis, which regulated the flavonoid biosynthesis in wax gourd. WGCNA was carried out to explore the correlation of the candidate gene with target genes in regulating the flavonoid pathway. To conclude, this study clarified the role of *BhiPRR6* gene in wax gourd fruit flavonoid biosynthesis and revealed its regulatory network. The findings will provide a theoretical basis of molecular genetic improvement in wax gourd.

Materials and methods

Plant materials, EMS mutagenesis, and mutant analysis

Wax gourd wild-type (WT) inbred line "BWT" was treated by 0.8% EMS to construct a mutant library, and a wax gourd mutant *hfc12* with high flavonoid content was generated. The 'BWT' and *hfc12* plants as parents to obtain an F_1 population and then were further self-crossing to establish the F_2 mapping population (486 individuals), including 361 'BWT' and 125 *hfc12* individuals, respectively. At the same time, F_1 and *hfc12* were backcrossed to obtain 238 individuals BC_1 population. All wax gourd varieties were planted in the Changsha experimental station (N 28°11′49″, E 112°58′429″) of the Hunan Vegetable Research Institute of Agricultural Science at the Apr. of 2018, Changsha, China. The flavonoid content, carotenoid content, chlorophyll a (Ca), and chlorophyll b (Cb) of the parents and F_2 population were determined at the same time. The flavonoid contents of wax gourd pericarp were determined by colorimetry using the Plant Flavonoids test kit (A142-1-1, Nanjing Jiancheng Bioengineering Institute). The carotenoid content, chlorophyll a (Ca), chlorophyll b (Cb) was determined by visible spectrophotometry using a Plant Carotenoid detection kit (BC4330, Beijing Solarbio Science & Technology Co., Ltd.). The segregation ratios in the F_2 and BC_1 populations were analyzed with the Chi-square (χ^2) goodness of fit test using R.

Transcriptome analysis

Total RNA was extracted from the frozen wax gourd pericarp of parents 'BWT' and *hfc12* collected at 40 days after pollination (DAP, 40DAP mature stage: Pericarp color change stage) were using TransZol Kit (TransGen Biotech, Inc., Beijing, China). The libraries (200–250 bp) of each sample were generated and sequenced on an Illumina HiSeq™ X-Ten platform for paired-end reads (BGI, Shenzhen, China). The quality of reads obtained was checked using the Fastqc program (Brown et al., 2017). Trimmmatic (v0.36) was used to trim the adapter sequences and remove low-quality reads (Bolger et al., 2014). The filtered reads were mapped to the wax gourd reference genome (*B. hispida* var. *B227*) (Xie et al., 2019) using the Salmon (v1.2.0) tool (Patro et al., 2017). TPM (Transcripts Per Million) was used to quantify the gene/transcript levels. DESeq2 was used to identify the DEGs with the criteria of $|\log2\text{Fold Change}| \geq 1$ and a false discovery rate (FDR) < 0.05 (Love et al., 2014). Gene Ontology (GO) was annotated using Plant Transcriptional Regulatory Map online software (http://plantregmap.gao-lab.org/), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was analyzed by Kobas3.0 online software. Both GO and KEGG were conducted using clusterProfiler (R, v3.12) with the corrected *P*-value ≤ 0.05 (Yu et al., 2012). The expression level of the genes in the whole tissue period were obtained from CuGenDB (http://cucurbitgenomics.org/) (Xie et al., 2019).

Metabolite profiling

Samples were extracted from the frozen wax gourd pericarp of 'BWT' and *hfc12* collected at 40 DAP with each three biological replicates. The biological samples were placed in a freeze dryer (Scientz-100F) for vacuum freeze drying. Using a grinder (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. Using ultra-performance liquid chromatography (UPLC; SHIMADZU Nexera X2) and tandem mass spectrometry (MS/MS) (Applied Biosystems 4500 QTRAP) to analyzed. A widely targeted metabolomics method was used based on the self-built database MWDB (Metware Biotechnology Co., Ltd. Wuhan, China) (http://www.metware.cn/). The metabolites were qualitatively analyzed based on the secondary spectrum information. The isotope signal is removed during analysis, including repeated signals of K^+, Na^+, $NH4^+$, and the repeated signals of fragment ions that are other larger molecular weight substances.

Unsupervised PCA was performed using the statistics function prcomp of the R statistical software (www.r-project.org). The hierarchical cluster analysis (HCA) results of samples and metabolites were presented as heatmaps with dendrograms. Pearson correlation coefficients (PCC; *r*) between the samples were calculated using the cor function in R and presented as a heatmap. Both HCA and PCC were carried out using the pheatmap (R, v1.0.12). For HCA, normalized signal intensities of metabolites (unit variance scaling) were visualized as a color spectrum. Significantly regulated metabolites between 'BWT' and *hfc12* were determined by VIP ≥1 and an absolute Fold change ≥ 2 or Fold change ≤ 0.5. VIP values were extracted from the orthogonal projections to latent structures discriminant analysis (OPLS-DA) result, with score plots and permutation plots, generated using MetaboAnalystR (R, v3.0) (Pang et al., 2020). The data were log-transformed (log_2) and mean-centered before OPLS-DA. A permutation test (200 permutations) was performed to avoid overfitting. Identified metabolites were annotated using KEGG Compound database (http://www.kegg.jp/kegg/compound/), annotated metabolites were then mapped to KEGG Pathway database (http://www.kegg.jp/kegg/pathway.html). Pathways with significantly regulated metabolites mapped to were then fed into metabolite sets enrichment analysis (MSEA).

Bulked segregant analysis RNA sequencing and linkage analysis

The DNA samples of the parental lines 'BWT' and *hfc12* were prepared as two pools for sequencing with each three biological replicates. From the F_2 population, 25 pericarps of wax gourd at 40 DAP for each plant with extremely high flavonoid content and 25 extremely low flavonoid content individuals were selected to construct two separation pools. In addition, DNA samples were extracted from individual plants in each pool and mixed in equimolar amounts to form wild-type mixed pools (low flavonoid content) and mutant mixed pools (high flavonoid content) for constructing genomic libraries. Four cDNA libraries (dominant mixed pool, recessive mixed pool, and two parent pools) were constructed using Truseq Nano DNA HT sample preparation kit (Illumina, USA). These four libraries were sequenced and evaluated on the Illumina HiSeq 4000 platform (Changsha, China) to generate 2×150 bp paired-end reads with an insert size of about 350 bp. The sequencing depth of the two mixed pools were 25×, and the sequencing depth of the two parent pools were 10×. Fastqc was used to evaluate the quality of the reads, and Trimmomatic to remove the adaptors and low-quality reads (Bolger et al., 2014; Brown et al., 2017). Hisat2 was used to map the clean reads to the wax gourd reference genome. SAMtools (v0.1.18) software was used to perform SNP calling with 90% loci depth coverage (Ramirez-Gonzalez et al., 2012). Further, Burrows-Wheeler Aligner (BWA) was used to align with the reference genome (*B. hispida* var. *B227*), and GATK (Genome analysis toolkit, v4.0) was used to detect SNPs and InDels. ANNOVAR was used to annotate the SNPs and InDels based on the GFF3 file of the wax gourd reference genome. In addition, the SNP-index algorithm was used to identify candidate regions of the genome related to flavonoid content, and calculated the difference in allele frequency between large-capacity pools (Takagi et al., 2013). Δ (SNP index) was the difference in SNP-index between the dominant mixed pool and the recessive mixed pool. The points with an SNP/InDel index less than 0.3 were filtered out of both pools. According to the SNP △(SNP-index) >0.5 and Euclidean distance (ED), the candidate regions related to the flavonoid content are selected with 95% confidence. Join Map software (v4.0) was used to construct the genetic linkage map with five SNPs as the window, and two SNPs as the step size.

Fine-mapping analysis

Mapping function was used to calculate the genetic map distance (cM) of the F_2 population 486 individuals from the recombination frequency (Voorrips, 2002). Based on the reference genome, Kompetitive Allele Specific PCR (KASP) was used to accurately determine SNPs and InDels biallelic genes at specific sites (Semagn et al., 2014). The CTAB method was used to extract DNA from all peel samples of 486 F_2 individuals for KASP analysis. The quality and quantity of the extracted DNA were verified by agarose gel electrophoresis and NanoDrop ND-2000 spectrophotometer, and the primers were designed by Primer5 (Table S4). The 200 bp sequence upstream and downstream of the SNPs were used to design the KASP primers, and the SNPs were typed based on the specific matching of primer end bases. KASP method and mix were to refer to PARMS SNP detection reagent manual (Gentides Biotech Co., Ltd (Wuhan) Synthesis), and were carried out on LightCycle® 96 Real-Time PCR System (Roche, Basel, Switzerland) in a 10μL reaction mixture. The selected genes and their primer sequences were listed in Table S4.

Real time fluorescence quantitative PCR

The qRT-PCR method was used to quantify the transcript abundance in the 40 DAP pericarp of the wax gourd plants. Reverse transcription was carried out using the HiScript®IIQ RT SuperMix for qRT-PCR (+gDNA wiper) (Vazyme Biotech Co. Ltd, Piscataway, NJ, United States). Primers were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/), with the PCR product size set at a range of 80–150 bp. The qRT-PCR was carried out on LightCycle® 96 Real-Time PCR System (Roche, Basel, Switzerland) in a 20μL reaction mixture by using ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech Co. Ltd,

Piscataway, NJ, United States). Three biological replicates and three technical replicates were maintained. *Actin* was used as the reference gene. The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression level of the genes. The selected genes and their primer sequences were listed in Table S4.

Correlation analysis

The WGCNA package in R language was used to construct a co-expression network and analyze the correlation of gene modules. RNA-seq FPKM data were used to perform WGCNA analysis (Langfelder and Horvath, 2008). The DEGs along DAFs with PCC ≥0.90 or ≤0.90 were selected for subsequent WGCNA with default parameters. Cytoscape was used to generate style and statistical a co-expression plot (Smoot et al., 2011). The DEGs in co-expression network were also subjected to GO and KEGG enrichment analysis, same method as above Transcriptome analysis of Materials and Methods part.

Results

Phenotypic characterization of 'BWT' and *hfc12*

EMS was used to mutate the WT wax gourd variety 'BWT' to construct a mutant library. After multiple generations of screening, a mutant *hfc12* plant with light-pericarp color was obtained. Phenotypic analysis revealed that the F_2 generation fruit peel of *hfc12* at 40 DAP was light green compared with the 'BWT' (Figure 1A). However, the *hfc12* color of the stem, leaves, and flowers was not significantly different from that of the 'BWT'; no significant difference was observed in the peel thickness also. The concentration of polyphenols and other compounds in the fruits of the *hfc12* and 'BWT' was analyzed. The flavonoid content (4.71 mg/g) in the *hfc12* was significantly higher than that of 'BWT' (2.85 mg/g). Besides, the content of chlorophyll a (5.52 mg/L) in the 'BWT' was significantly higher than that of *hfc12* (3.38 mg/L), and the content of chlorophyll b (3.83 mg/L) in the 'BWT' was significantly higher than that of *hfc12* (1.65 mg/L). No significant difference was observed between the *hfc12* and the 'BWT' in carotenoid concentration and content (Figure 1B).

Identification and functional analysis of DEGs

To further understand the flavonoid response patterns of 'BWT' and *hfc12* at the transcriptional level, genes related to higher flavonoid levels were identified from wax gourd. Transcriptome sequencing (RNA-seq) was performed using the *hfc12* and 'BWT' fruit frozen pericarp samples collected at 40 DAP. After deleting the adaptors and low-quality reads from the raw data, 55.05 Gb clean data were generated from the six samples. The number of reads per sample ranged from 20.27 million to 28.85 million, with an average of 22.80 million (Table S1). Approximately 95% of the clean reads were mapped to the wax gourd (*B. hispida* cv. B227) reference genome. Furthermore, the correlation coefficient and principal component analysis (PCA) suggested a high degree of similarity between the biological repeats (Figure 2A). A total of 828 DEGs were identified between 'BWT' and *hfc12* at 40 DAP, including 286 upregulated and 542 downregulated genes (Figure 2B).

GO analysis was performed to identify the primary biological functions of all the DEGs between 'BWT' and *hfc12* with the corrected *P*-value < 0.05. All DEGs were assigned GO terms of the three major categories: cellular component, molecular function, and biological process (Figure 2C). DEGs were primarily assigned in cell part, cell, organelle, and membrane in the cellular component category. In the molecular function category, catalytic activity and binding were significantly assigned. In the biological process category, 56% of DEGs were primarily assigned in the metabolic processes (GO: 0008152), specially. Cellular process, response to stimulus, biological regulation, and regulation of biological process were also highly assigned in the biological process category.

Furthermore, KEGG pathway analysis revealed that DEGs were associated with 32 common metabolic and biological pathways (Figure 2D). DEGs were preferably enriched in the metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction, protein processing in endoplasmic reticulum. Besides, DEGs were significantly enriched ($-\log_{10}$(Corrected_*P*_Value) >0.1) in sesquiterpenoid and triterpenoid biosynthesis, and brassinosteroid biosynthesis Photosynthesis, MAPK signaling pathway - plant, fructose and mannose metabolism, and carbon fixation in photosynthetic organisms were also enriched.

Identification and enrichment analysis of DAFs in the fruits

TRAP-MS/MS was performed to compare the differences in flavonoid composition between the 'BWT' and *hfc12* wax gourd fruit frozen pericarp samples collected at 40 DAP (Figure S1A). A total of 138 flavonoids were detected in all samples, including 29 flavonoid carbonosides, 26 flavonols, 51 flavonoids, 13 dihydroflavones, two chalcones, and eight isoflavones, eight kinds of tannins and one kind of proanthocyanidins (Table S2). Among them, a total of 31 DAFs were identified between 'BWT' and *hfc12* (fold change ≥ 2 or fold change ≤ 0.5, and variable importance in projection (VIP) ≥1), of which 28 were upregulated and three were downregulated. These DAFs were either flavonoids or flavonoid carboglycosides. Twelve

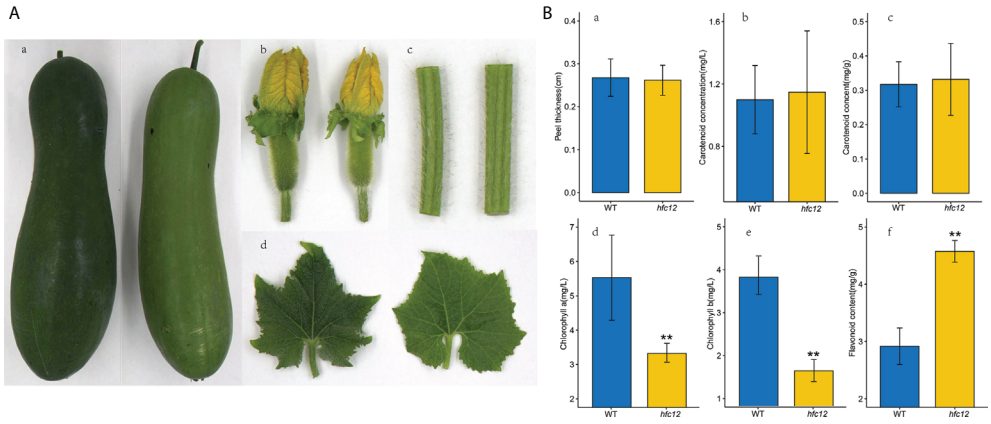


FIGURE 1
Phenotypic characterization and physiological indicators of WT and *hfc12*. **(A)** Phenotypic characterization of fruits at 40 DAP (a), flowers (b), stems (c), and leaves (d) of WT (left) and *hfc12* (right). **(B)** Physiological indicators including peel thickness (a), carotenoid concentration (b), carotenoid content (c), chlorophyll a content (d), chlorophyll b content (e), and flavonoid content (f) of 'BWT' (left) and *hfc12* (right). Asterisks indicate significance according to the t-test (**P< 0.01).

flavonoids were identified in DAFs (Figure 3A). Besides, one chalcone, two flavanols, three flavonols, three isoflavones, three dihydroflavonol, and seven flavonoid carbonosides were also among the DAFs. Of which, chrysoeriol-7-O-(6''-acetyl) glucoside and isorhamnetin-3-O-(6''-acetylglucoside) were the highest upregulated in the *hfc12*, with a $\log_2$FC value of 11.09 and 10.72, respectively. Orientin-7-O-glucoside, chrysoeriol-7-O-(6''-malonyl) glucoside, and quercetin-3-O-(4''-O-glucosyl) rhamnoside were the three downregulated metabolites in the *hfc12*, with a $\log_2$FC value of -1.18, -1.07, and -1.06, respectively (Figure 3B).

Furthermore, KEGG enrichment analysis revealed that the DAFs were enriched with flavone and flavonol biosynthesis (ko00944), flavonoid biosynthesis (ko00941), isoflavonoid biosynthesis (ko00943), and secondary metabolite biosynthesis pathways (ko01110). These observations indicated that the *hfc12*

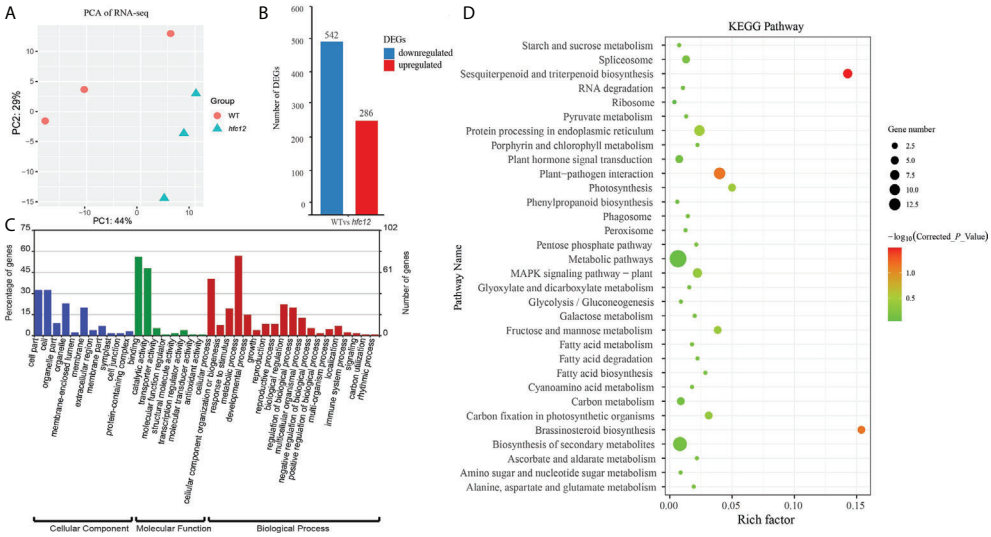


FIGURE 2
RNA-seq transcriptome analysis of WT and *hfc12*. **(A)** PCA plot of DEGs between 'BWT' and *hfc12*. **(B)** Differentially expressed genes (DEGs) at 40 DAP of WT and *hfc12*. **(C)** GO enrichment analysis of DEGs. Blue bars for cellular component, green bars for molecular function, and red bars for biological process. **(D)** KEGG analysis of DEGs.

10.3389/fpls.2022.1019787

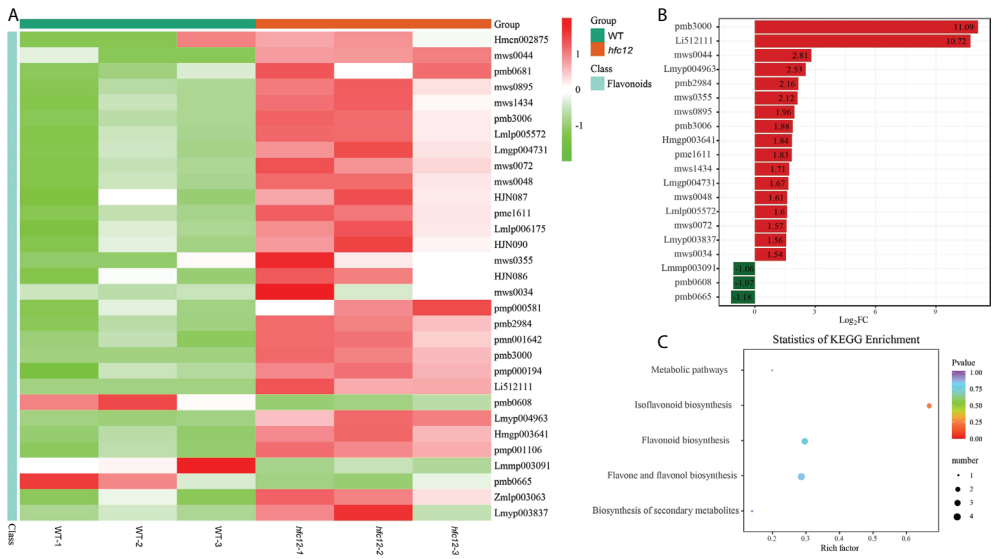


FIGURE 3
Metabonomic analysis of WT and *hfc12*. **(A)** Heatmap of DAFs between WT and *hfc12*. The columns represented samples, and the rows represented the differential metabolites. The differential metabolite cluster tree was shown on the left side of the plot. The different colors indicated the values obtained after standardization of the relative content (red represents high content, green represents low content). **(B)** Bar chart of DAFs. **(C)** KEGG enrichment scatter plot of DAFs.

plant have mutations in the flavonoid metabolism pathway, probably affecting fruit development and peel color changes (Figure 3C).

Mapping of wax gourd *hfc12*

The F_2 (n = 486) population isolated from the cross between the WT 'BWT' and *hfc12* mutant the was used for genetic analysis and primary mapping. Genetic population analysis and chi-square test revealed that the low flavonoid content (361 individuals, dark-green pericarp) and the high flavonoid content (125 individuals, light-green pericarp) traits of the F_2 generation were separated at a ratio of 3:1, and that of the backcross population at a ratio of 1:1 (Table 1), consistent with the Mendelian independent genetic rules. These observations indicated that a single recessive nuclear gene controls the trait related to high flavonoid content.

Four parental DNA pools (dominant pool, recessive pool, dominant progeny pool, and recessive progeny pool) were constructed, and the WT pool and mutant pool were mixed to construct a genomic library for BSA-seq. Genome-wide BSA-seq using these DNA pools resulted in 67 Gb of data. The two parental lines generated 0.76 million paired-end reads of 'BWT' and 0.61 million paired-end reads of *hfc12*. The sequencing depths were 10.47× and 10.05×, and the genome coverage rates were 95.60% and 95.93% in the 'BWT' and *hfc12*, respectively. Similarly, the alignment of the dominant sexual progeny pool and recessive progeny pool were 20.01× and 18.92× of sequencing depth, and genome coverage of 95.67% and 95.93%, respectively (Table S3). Statistical analysis suggested that the sequencing data were reliable for BSA-seq analysis. A total of 21,548 SNPs were identified between the 'BWT' and the *hfc12* libraries (base quality value ≥ 20, mapping quality value ≥ 20, base depth (two F_2 mixed pools) ≥ 2 and ≤ 60, and base depth (parents) ≥ 2 and ≤ 60). There were two regions on the whole

TABLE 1 Genetic analysis and flavonoids content of wax gourd *hfc12*.

Type	Population	WT	*hfc12*	Theoretical ratio	Actual ratio	χ^2	Dominant/ Recessive	WT flavonoids range (mg/g)	*hfc12* flavonoids range (mg/g)	WT flavonoids mean (mg/g)	*hfc12* flavonoids mean (mg/g)
F_2	486	361	125	3:1	2.9:1	0.13	R	2.36-3.28	3.97-5.84	2.73	4.78
BC_1	238	116	122	1:1	0.95:1	0.15	R	-	-	-	-

$\chi_c^2 < \chi^2_{0.05(1)}=3.84$, $P>0.05$.

genome of the recessive pool with Δ (SNP index) values above 0.5 and close to 1, which were located in the 25-30 Mb range of chromosome 12, and a total of 456 candidate sites for the target region have been obtained.

Fine-mapping and candidate gene analysis

All F_2 individuals from 'BWT' × *hfc12* were used for KASP molecular markers for genotyping. Based on the genotypes and phenotypes, a total of 78 recombinant plants were identified for further population location. Finally, fine-mapping places the *hfc12* locus in the genomic region, flanked by markers 24.16 Mb (SNP12G24166107) to 35.90 Mb (SNP12G35908349), a 11.7Mb on chromosome 12 (Figure 4A). According to the wax gourd reference genome '*Benincasa hispida* var. B227' database, 11 predicted genes were annotated in the candidate region, including nine intron mutations, two 5' UTR mutations, one 3' UTR mutation, and one coding sequence (CDS) mutation (Table S5). Nine polymorphic SNP markers were further developed from 24.16 Mb to 35.90 Mb region, and the mutation site was finally restricted to the 452.366 kb region (SNP12G25509139 to SNP12G25961505) on chromosome 12 by

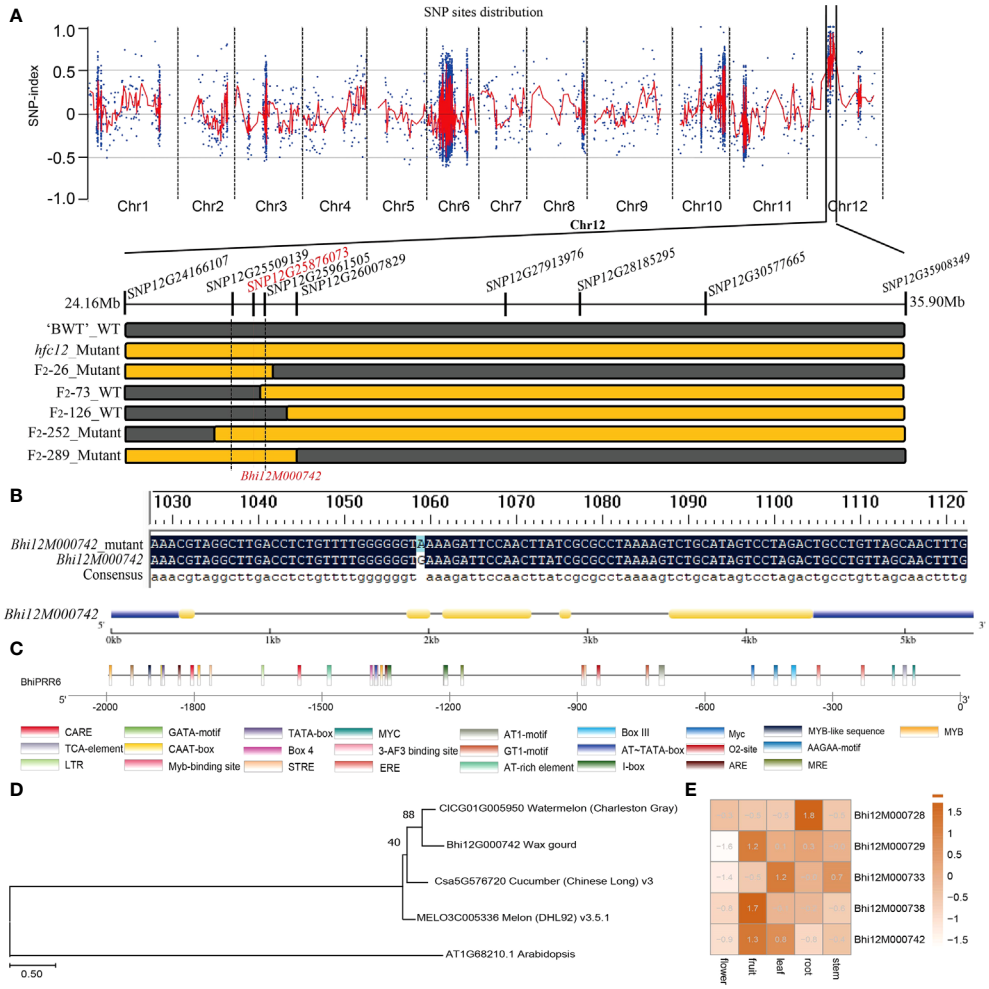


FIGURE 4
Cosegregation analysis and genetic variation in *BhiPRR4* of wax gourd *hfc12*. **(A)** Gene location and mutation site in wax gourd. BSA-seq generated the SNP distribution Δ(SNP index)] plot; the vertical axis represents the chromosome, and the red vertical line represents the location of the *hfc12* SNP. The highest point shown is the locus 25 Mb to 30 Mb on chromosome 12. Based on F_2 plants with extreme phenotypes, the genetic map was drawn on the 452.366 kb region (SNP12G25509139 to *SNP12G25961505*). **(B)** *Bhi12G000742* (*BhiPRR6*) mutation of C to T in the SNP on the positive strand (G to A on the negative strand) of the second exon (red line). The yellow rectangle and the black line represent the exons and introns, respectively. **(C)** Schematic representation of *BhiPRR6* promoter 2000 bp upstream of the CDS region. *Cis*-elements are shown in rectangles with different colors. **(D)** Multiple sequence alignment of *BhiPRR6* proteins with watermelon, melon, cucumber, and *Arabidopsis*. **(E)** Different tissue expression levels of genes in the candidate interval.

multiple KASP molecular markers. Within this interval, there were a total of 219 SNPs, and involving 5 genes (Figure 4E). However, only one SNP (SNP12G25876073) is located in the gene CDS region and co-segregated with the phenotypes among the F_2 population, resulting in non-synonymous mutations between WT and *hfc12* (Figure 4B). PCR amplification and sequencing identified the SNP-25876073 was located in the second exon of *Bhi12G000742* (*BhiPRR6*). A C-T mutation was detected at 1060 bp of mutant *hfc12* CDS region; CTT was mutated to TTT (+), and Glu (E) to Lys (K) (-) (Figure 4B).

Annotation and protein homology comparison found that the candidate gene was a typical two-component response regulator PRR family gene. The analysis identified the gene as *BhiPRR6*, 1802 bp long with five exons and four introns. It had theoretical pI (isoelectric point) and Mw (molecular weight) of 6.88 and 67180.10, respectively, and was located on chromosome 12 (25876073–25877875). The promoter region (2000 bp upstream) of *BhiPRR6* had 26 *cis*-elements, including typical ARE, AT-rich element, ethylene-responsive element (ERE), and salicylic acid-responsive element (TCA-element). It also included various typical MYB *cis*-acting elements such as MRE, MYB, MYB-like sequence, MYC, and MYB-binding site (Figure 4C). Besides, *BhiPRR6* was highly homologous to the watermelon (*Citrullus lanatus* subsp. *vulgaris* cv. 97103) *CLCG01G005950*, the muskmelon (*Cucumis melo* L. cv. DHL92) *MELON3C005336* (*CmPRR5*, *LOC103504590*), the cucumber (*Cucumis sativus* L. var. *sativus* cv. 9930) *Csa5G576720*, and the *Arabidopsis AT1G68210* (*APRR6*). The closest homology was found between *BhiPRR6* and the watermelon *CLCG01G005950* gene, with a bootstrap value of 89 (Figure 4D). Analysis of the expression levels of five candidate genes in the KASP regions showed that *BhiPRR6* was highly expressed in fruits and leaves (Figure 4E).

Correlation analysis of *BhiPRR6* in flavonoids biosynthetic pathways

WGCNA was performed using RNA-seq data to study the correlation pattern among the genes in wax gourd. All 971 co-expressed genes with *BhiPRR6* as the hub gene were clustered into eight main modules (marked in different colors) (Table S6). Genes in the same module were highly related and co-expressed. The turquoise module had the maximum number of DEGs (258), followed by the blue (148), brown (145), and yellow (144) modules (Figure S1D). Co-expression analysis suggested a high correlation between 78 genes and *BhiPRR6*, with a strict edge weight (topological overlap matrix, TOM ≥ 0.30). Among them, F-box family protein (*Bhi11G001950*), glycerol-3-phosphate acyltransferase 1 (GPAT1, *Bhi05G001763*), acyl carrier protein 4 (ACP4, *Bhi02G000477*), and CBL-interacting protein kinase 1 (CIPK1, *Bhi12G001773*) showed a high degree of connectivity, indicating certain biological functions. *Bhi10G001402*, *Bhi06G000135*, *Bhi09G001849*, and *Bhi05G000810* showed the highest co-expression with the hub gene *BhiPRR6* (Figure 5A). In particular, co-expressed genes were related to flavonoid biosynthesis, including a structural gene 4CL (4-coumarate-CoA ligase, *Bhi11G001762*), with a TOM value of 0.35. UDP-glucose: flavonoid-3-O-glycosyltransferase (UFGT; *Bhi09G002071*, *Bhi03G000738*, *Bhi05G000029*, and *Bhi05G000624*) and photosystem I chlorophyll a/b-binding protein 6 (LHCA6, *Bhi02G001494*) were also highly co-expressed with *BhiPRR6*. An MYB1R1 (*Bhi11G000031*) also showed a high correlation with *BhiPRR6* involved flavonoid synthesis. A WD40 protein (*Bhi12G000575*) and a 4CL structure protein (*Bhi11G001234*) were also present in this network.

Furthermore, the expression levels of the co-expressed genes were analyzed using RNA-seq data. The analysis divided 79

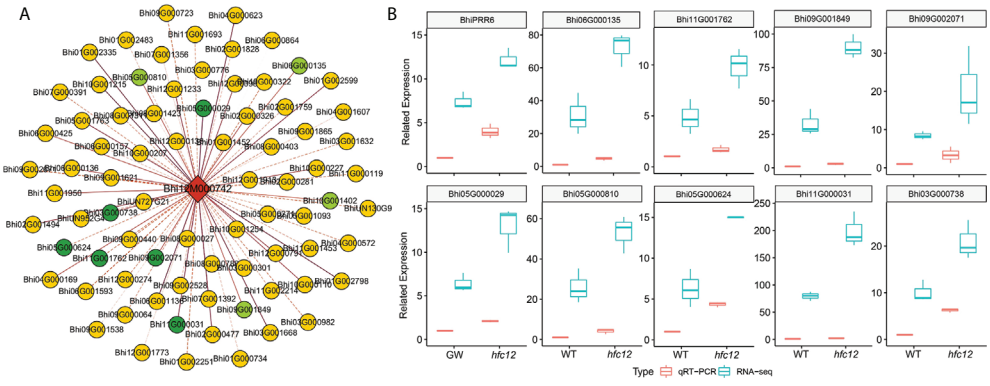


FIGURE 5
Co-expression regulatory network and expression of *BhiPRR6* and other interacting genes in wax gourd. **(A)** Co-expression network analysis with *BhiPRR6* as hub gene. The light-green circles represent the genes with the highest correlation (Topological overlap matrix, TOM > 0.52). The dark-green circles represent the candidate genes that may regulate the synthesis of flavonoids with *BhiPRR6* (Table S5). **(B)** The expression pattern of *BhiPRR6* and ten co-expressed genes by both Quantitative real-time PCR (qRT-PCR) and RNA-seq analysis.

genes (hub gene and 78 co-expressed genes) into two categories based on the expression pattern. Among them, most co-expressed genes (74 genes) were similar to *BhiPRR6*, with expression levels significantly high at 40 DAP of *hfc12*. In the network, ten candidate co-expressed gene annotations may be related to the flavonoid pathway, with the similar expression pattern of *BhiPRR6* (Figure 5A). RNA-seq and qRT-PCR were performed to analyze the transcript level of *BhiPRR6* and the co-expressed genes in the mutant *hfc12* and the WT 'BWT' parent individuals. As shown in Figure 5B, the expression level of *BhiPRR6* was highest in *hfc12*, two times more than that in WT. At the same time, ten co-expressed genes closely related to *BhiPPR6* were basically the same as their expression forms, and they were all highly expressed in the mutant *hfc12*. Of which, the highest expression was the *Bhi09G001849* gene, and its expression in the *hfc12* was about four times that of the 'BWT'. The qRT-PCR results were consistent with the RNA-seq results, with an R^2 above 0.80.

Discussion

Wax gourd is a globally preferred vegetable due to its high nutritional value and health benefits. Among the different bioactive components, flavonoids, such as anthocyanins, flavonols, and flavan-3-ols, in the fruit have antiangiogenetic and anticancer properties (Jiang et al., 2015). Much effort has been made to elucidate the flavonoid biosynthetic pathway and identify the regulatory factors. A previous study showed that grape (*Vitis vinifera* L.) adapted to intense light by increasing the expression of flavonoid biosynthetic genes in the skin, leading to increased anthocyanin, proanthocyanidin, and flavonol content (Dobre et al., 2014). In litchi (*Litchi chinensis*), fruit bagging inhibited anthocyanin accumulation and biosynthetic genes encoding CHS, CHI, F3H, dihydroflavonol 4-reductase (DFR), anthocyanin synthase (ANS), and UFGT (Koyama et al., 2012). In this study, the flavonoid content was high in the EMS-induced mutant *hfc12*, indicating the existence of regulatory factors for flavonoid content in wax gourd. Metabolites, including flavonoid carbonosides, flavonols, isoflavones, flavanols, dihydroflavonol, and chalcones, were found to be differentially accumulated in the *hfc12*. Detailed analysis revealed that the DAFs and DEGs between the 'BWT' and *hfc12* were significantly enriched in the flavonoid pathway (Figure 2, Figure 3). These observations collectively suggested that various DAFs and DEGs have related roles, and transcription factors may potentially coordinate the synthesis of flavonoids in wax gourd.

Fine-mapping identified a typical two-component system (TCS) PRR transcription factor *BhiPRR6* on chromosome 12 of wax gourd, with a C-T nonsynonymous mutation in the positive chain CDS region (+), in the *hfc12* with higher levels of flavonoids. The candidate gene *BhiPRR6* was highly homologous to *APRR6* of *Arabidopsis*. The TCS signaling is an important mechanism regulating various pathways in prokaryotes and eukaryotes (Wei et al., 2011). This pathway usually consists of three signal elements: histidine kinases (HKs), response regulators (RRs), and histidine phosphate transfers (HPs) (Stock et al., 2000). The HK protein could be got phosphorylated, and this phosphate group gets continuously transferred to the conserved Asp residues in the RR receptor domain (Hwang et al., 2002). Phosphorylated RR protein directly or indirectly regulates the activity of downstream genes. The HP protein acts as a bridge in the phosphate transfer between HK and RR (Grefen and Harter, 2004). These response modifiers are divided into three subfamilies: type-A RRs with only receptor domains, type-B RRs with receptor domains fused to DNA binding domains, and PRR with atypical receptor domains (Huo et al., 2020). Type-B RRs bind to multiple cis-elements in the promoters of the target genes, such as mitogen-activated protein kinase (MAPK), or other regulatory genes, thereby participating in diverse regulatory responses (Wurgler-Murphy and Saito, 1997). The members of the PRR family have significant sequence similarity with the RRs in the putative receptor domain. They do not have a conserved D-D-K motif in the receptor domain and lack conserved Asp residues but can be used as the final product of two-component phosphorylation in plants. In this study, as Figure 1 and Figure S1B shown, *BhiPRR6* was the highest expressed in fruit, and the flavonoid content of the *hfc12* was significantly higher than that of the WT, indicating the importance of *BhiPRR6* in regulating flavonoid biosynthesis in wax gourd.

Several factors, such as light, regulate flavonoid synthesis. Higher plants use sensory photoreceptors to accurately perceive light from UV-B to far-red wavelengths and chlorophyll and carotenoids of the photocomplex in photosynthesis (Casal, 2013). One of the most important is the superfamily of plant pigments, including photoreceptors (PHYA, PHYB, PHYC, PHYD, PHYE) that absorb red/far-red light and leuco pigments (CRY1, CRY2, CRY3) and photosensitive proteins (PHOT1, PHOTO2) (Wagner et al., 2005). In this study, the flavonoid content of the *hfc12* significantly increased, accompanied by a substantially lower chlorophyll a and chlorophyll b, indicating a close association between flavonoid synthesis and light pigments (Figure 1). Various transcription factors through differential expression regulate the biosynthesis of different flavonoids in response to specific light wavelengths (Zoratti et al., 2014). Moreover, in higher plants, the light pigments may be a His protein kinase (HKs). The light signals use the light-regulated phosphorous transfer mechanism, probably using a separate receptor protein RR as an intermediate (Huq et al., 2010). Therefore, HPs and RRs play essential roles in light signaling (Yeh et al., 1997). In *Arabidopsis*, *ARR4* transcription factor

acts as a signaling module in the light signal transduction pathway (Haberer and Kieber, 2002). A study using GBS-based BSA-seq analysis in pepper identified a candidate gene *CaPRR2* related to plastid development and pigment biosynthesis, affecting the levels of chlorophyll a and chlorophyll b and controlling the color of fruit (Lee et al., 2020).

Furthermore, a typical MYB DNA-binding domain was identified in the candidate gene *BhiPRR6* sequence, and the mutation occurred in this region, indicating it has a function was similar with the MYB transcription factor. Studies have shown that MYB transcription factors coordinate and regulate flavonoid structural genes by activating or inhibiting their expression. MYB transcription factors related to flavonoid biosynthesis have been identified in various plants, with few responsive to light. In mango (*Mangifera indica* L.), *MiMYB1* at the transcript level regulates the light-dependent red coloration of fruit skin (Kanzaki et al., 2020). Moreover, the MYB transcription factor directly interacts with the MYB recognition element (MRE) in the promoter region of the structural genes (Yao et al., 2017). Studies have also reported that MRE is essential for light-induced expression of structural flavonoid genes, such as *CHS* (Takos et al., 2006). Typical MRE *cis*-elements were found in the promoter region of the candidate gene *BhiPRR6*. Other MYB typical *cis*-elements also were found in the promoter region of *BhiPRR6*, such as MRE, MYB, MYB-like sequence, MYC, MYB-binding site, and Myc. Subsequent structural analysis revealed the typical TCS and MYB-like functions of *BhiPRR6* in regulating the light-regulated pathway, leading to changes in flavonoid content of wax gourd.

Subsequently, a co-expression network was constructed using BhiPRR6 as hub gene to find why it can regulate the synthesis of flavonoids in wax gourd and leads to the change of peel color (Figure 5). Interestingly, many flavonoid-related structural genes were identified as the co-expressed genes, such as *Bhi4CL* and *BhiUFGT*; *4CL* forms the backbone unit for flavonoids, and *UFGT* acts in the last step of flavonoid biosynthesis for glycosylation modification (Jaakola, 2013). The co-expression plot also indicated the co-expression of *6LHCA6*, *MYB1R1*, and *WD40* with the hub gene *BhiPRR6*. Among them, MYB-bHLH-WD40 ternary complex (MBW complex) is a typical representative to positively regulate the biosynthesis of anthocyanins and flavonoids (Xu et al., 2021). The expression pattern of the co-expressed genes showed consistency with that of *BhiPRR6* and its associated target genes, indicating that they are similarly regulated in this *hfc12* (Figure S2). These observations in the *hfc12* suggest that *BhiPRR6* binds to the co-expressed genes through various *cis*-elements on the promoter. The MYB DNA-binding domain of *BhiPRR6* may regulate the conduction of light signals and the function of downstream genes and thereby inhibit flavonoid biosynthesis in wax gourd. In summary, it is collectively indicated that *BhiPRR6* and the co-expressed genes are highly expressed during the later stages of fruit development, inhibiting the transmission of light signals and leading to the reduction in flavonoids, which in turn leads to changes in peel color. This study provided new insights into the synthesis and regulation of flavonoids in wax gourd fruits and laid the foundation for breeding wax gourd varieties with higher levels of flavonoids.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI BioProject: PRJNA781301.

Author contributions

LX, JW and FL: conceptualization, investigation, writing – review and editing, and preparation. JW: formal analysis, writing – original draft, and writing – original draft. BM, YC, LP, WX, and YL: investigation, formal analysis, and methodology. XS and CX: conceptualization, resources, supervision, funding acquisition, project administration, and writing – review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was financed by National Natural Science Foundation of China (Grant No.32102398). The authors are grateful to Professor Xuexiao Zou from Hunan Agricultural University for providing the experimental environment and critical comments of this manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1019787/full#supplementary-material

SUPPLEMENTARY FIGURE 1

(A). PCA plot of Metabonomic analysis. **(B)**. Transcript levels of *BhiPRR6* in flower, fruit, leaf, root, and stem of wax gourd. **(A)**. RNA-seq analysis. **(C)**. Soft threshold filter plot. The vertical axis on the left represented the square of the correlation coefficient between log(k) and log(p(k)) in the corresponding network. The higher the square of the correlation coefficient, the closer the network was to the distribution without network scale. The vertical axis of the right figure represented the mean value of all gene adjacency functions in the corresponding gene module. This study was used 7 as the soft threshold. **(D)**. Co-express network-level cluster analysis. The branches and different colors of the clustering tree are used to indicate different gene modules. **(E)**. GO enrichment analysis of co-expressed genes. Blue bars for cellular component, green bars for molecular function, and red bars for biological process. **(F)**. KEGG analysis of co-expressed genes.

SUPPLEMENTARY FIGURE 2

The expression heatmap of *BhiPRR6* (red color) and all highly co-expressed genes by RNA-seq analysis. All samples were collected at 40 DAP peels of wax gourd.

References

Almeida, J., D'Amico., E., Preuss, A., F. Carbone. Vos, C., Deiml, B., Mourgues, F., et al. (2007). Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria xananassa*). *Arch. Biochem. Biophys.* 465, 61–71. doi: 10.1016/j.abb.2007.04.040

Azuma, A., Hiroshi, Y., Yoshiko, K., and Shozo, K. (2012). Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. *Planta* 236, 1067–1080. doi: 10.1007/s00425-012-1650-x

Baranac, J. M., Petranovi, N. A., and Dimitri-Markovi, J. M. (1997). Spectrophotometric study of anthocyan copigmentation reactions. *2. Malvin nonglycosidized flavone quercetin. J. Agr. Food Chem.* 45, 1698–1700. doi: 10.1021/jf9606114

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Usadel, trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brown, J., Meg, P., and Lee, A. M. C. (2017). McCue, FQC dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33, 3137–3139. doi: 10.1093/bioinformatics/btx373

Casal, J. J. (2013). Photoreceptor signaling networks in plant responses to shade. *Ann. Rev. Plant Biol.* 64, 403. doi: 10.1146/annurev-arplant-050312-120221

Chen, C., Zhou, G., Chen, J., Liu, X., Lu, X., Chen, H., et al. (2021). Integrated metabolome and transcriptome analysis unveils novel pathway involved in the formation of yellow peel in cucumber. *Int. J. Mol. Sci.* 22, 1494. doi: 10.3390/ijms22031494

Dobre, C., Toma, E., Lucian, I., and Lavinia, M. B. (2014). *Benincasa hispida* (wax gourd) technological characteristics and biological proprieties. *Rom. Biotech. Lett.* 19, 9504–9509.

Feder, A., Burger, J., Gao, S., Lewinsohn, E., Katzir, N., Schaffer, A. A., et al. (2015). A kelch domain-containing f-box coding gene negatively regulates flavonoid accumulation in muskmelon. *Plant Physiol.* 169, 1714–1726. doi: 10.1104/pp.15.01008

Grefen, C., and Harter, K. (2004). Plant two-component systems: Principles, functions, complexity and cross talk. *Planta* 219, 733–742. doi: 10.2307/23388587

Haberer, G., and Kieber, J. J. (2002). New insights into a classic phytohormone. *Plant Physiol.* 128, 354–362. doi: 10.1104/pp.010773

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465. doi: 10.1093/jxb/erq442

Hu, D. G., Li, Y. Y., Zhang, Q. Y., Li, M., Sun, C. H., Yu, J. Q., et al. (2017). The R2R3-MYB transcription factor *MdMYB73* is involved in malate accumulation and vacuolar acidification in apple. *Plant J.* 91, 443–454. doi: 10.1111/tpj.13579

Huo, R., Liu, Z. N., Yu, X. L., and Li, Z. Y. (2020). The interaction network and signaling specificity of two-component system in *Arabidopsis*. *Int. J. Mol. Sci.* 21 (14), 4898. doi: 10.3390/ijms21144898

Huq, E., Al-Sady, B., and Quail, P. H. (2010). Nuclear translocation of the photoreceptor phytochrome b is necessary for its biological function in seedling photomorphogenesis. *Plant J.* 35, 660–664. doi: 10.1046/j.1365-313X.2003.01836.x

Hwang, I., Chen, H. C., and Sheen, J. (2002). Two-component signal transduction pathways in *Arabidopsis*. *Plant Physiol.* 129, 500–515. doi: 10.1104/pp.005504

Imran, M., Abdur, R., and Abu-Izneid, T. (2019). Luteolin, a flavonoid, as an anticancer agent: A review. *BioMed. Pharmacother.* 112, 108612. doi: 10.1016/j.biopha.2019.108612

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18 (9), 477–483. doi: 10.1016/j.tplants.2013.06.003

Jaakola, L., and Hohtola, A. (2010). Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ.* 33, 1239–1247. doi: 10.1111/j.1365-3040.2010.02154.x

Jaakola, L., Poole, M., and Jones, M. O. (2019). A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiol.* 153, 1619–1629. doi: 10.1104/pp.110.158279

Jiang, B., Liu, W. R., Xie, D. S., Peng, Q. W., He, X. M., Lin, Y. E., et al. (2015). High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing. *BMC Genomics* 16, 1035. doi: 10.1186/s12864-015-2220-y

Kanzaki, S., Ichihi, A., Tanaka, Y., Fujishige, S., and Shimizu, K. (2020). The R2R3-MYB transcription factor *MiMYB1* regulates light dependent red coloration of 'Irwin' mango fruit skin. *Scientia Hortic.* 272, 109567. doi: 10.1016/j.scienta.2020.109567

Koyama, K., Ikeda, H., Poudel, P. R., and Goto-Yamamoto, N. (2012). Light quality affects flavonoid biosynthesis in young berries of Cabernet sauvignon grape. *Phytochemistry* 78, 54–64. doi: 10.1016/j.phytochem.2012.02.026

Kurlovs, A. H., Snoeck, S., Kosterlitz, O., Leeuwen, T., and Clark, R. M. (2019). Trait mapping in diverse arthropods by bulked segregant analysis. *Curr. Opin. Insect Sci.* 36, 57–65. doi: 10.1016/j.cois.2019.08.004

Langfelder, P., and Horvath, S. (2008). WGCNA: An r package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Lee, S. B., Kim, J. E., Kim, H. T., Lee, J. M., Kim, B. S., and Lee, J. M. (2020). Genetic mapping of the c1 locus by GBS-based BSA-seq revealed pseudo-response regulator 2 as a candidate gene controlling pepper fruit color. *Theor. Appl. Genet.* 133, 1897–1910. doi: 10.1007/s00122-020-03565-5

Li, Z., Peng, R., and Yao, Q. (2021). *SlMYB14* promotes flavonoids accumulation and confers higher tolerance to 2,4,6-trichlorophenol in tomato. *Plant Sci.* 303, 110796. doi: 10.1016/j.plantsci.2020.110796

Li, Y., Wen, C., and Weng, Y. (2013). Fine mapping of the pleiotropic locus b for black spine and orange mature fruit color in cucumber identifies a 50 kb region containing a *R2R3-MYB* transcription factor. *Theor. Appl. Gene1.* 26 (8), 2187–2196. doi: 10.1007/s00122-013-2128-3

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

Martens, S., Preuss, A., and Matern, U. (2010). Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* l. *Phytochemistry* 71, 1040–1049. doi: 10.1016/j.phytochem.2010.04.016

Pang, Z. Q., Chong, J., Li, S. Z., and Xia, J. G. (2020). MetaboAnalystR 3.0: Toward an optimized workflow for global metabolomics. *Metabolites* 10 (5), 186. doi: 10.3390/metabo10050186

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Premathilake, A. T., Ni, J. B., Bai, S. L., Tao, R. Y., Ahmad, M., and Teng, Y. W. (2020). R2R3-MYB transcription factor *PpMYB17* positively regulates flavonoid biosynthesis in pear fruit. *Planta* 252 (4), 59. doi: 10.1007/s00425-020-03473-4

Ramirez-Gonzalez, R. H., Bonnal, R., Caccamo, M., and Maclean, D. (2012). Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments. *Source Code Biol. Med.* 7. doi: 10.1186/1751-0473-7-6

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Shi, Q., Du, J., Zhu, D., Li, X., and Li, X. (2020). Metabolomic and transcriptomic analyses of anthocyanin biosynthesis mechanisms in the color mutant *Ziziphus jujuba* cv. *Tailihong*. *J. Agr. Food Chem.* 68, 15186–15198. doi: 10.1021/acs.jafc.0c05334

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432. doi: 10.1093/bioinformatics/btq675

Sreenivas, K. M., Chaudhari, K., and Lele, S. S. (2011). Ash gourd peel wax: Extraction, characterization, and application as an edible coat for fruits. *Food Sci. Biotechnol.* 20, 383–387. doi: 10.1007/s10068-011-0054-1

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Annu. Rev. Biochem.* 69, 183–215. doi: 10.1146/annurev.biochem.69.1.183

Sudheeran, P. K., Sela, N., Carmeli-Weissberg, M., Ovadia, R., and Alkan, N. (2021). Induced defense response in red mango fruit against colletotrichum gloeosporioides. *Hortic. Res.* 8, 17. doi: 10.1038/s41438-020-00452-4

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., and Mitsuoka, C. (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

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142, 1216–1232. doi: 10.1104/pp.106.088104

Tomás-Barberán, F. A., Gil, M. I., Cremin, P., Waterhouse, A. L., Hess-Pierce, B., and Kader, A. A. (2001). HPLC–DAD–ESIMS analysis of phenolic compounds in nectarines, peaches, and plums. *J. Agr. Food Chem.* 49, 4748–4760. doi: 10.1021/jf0104681

Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wagner, J. R., Brunzelle, J. S., Forest, K. T., and Vierstra, R. D. (2005). A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 438, 325–331. doi: 10.1038/nature04118

Wei, Y. Z., Hu, F. C., Hu, G. B., Li, X. J., Huang, X. M., Wang, H. C., et al. (2011). Differential expression of anthocyanin biosynthetic genes in relation to anthocyanin accumulation in the pericarp of litchi chinensis sonn. *PloS One* 6, e19455. doi: 10.1371/journal.pone.0019455

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Bio* 5, 218–223. doi: 10.1016/S1369-5266(02)00256-X

Wurgler-Murphy, S. M., and Saito, H. (1997). Two-component signal transducers and MAPK cascades. *Trends Biochem. Sci.* 22, 172–176. doi: 10.1016/S0968-0004(97)01036-0

Xie, D., Xu, Y., Wang, J., Liu, W., and Zhang, Z. (2019). The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. *Nat. Commun.* 10, 5158. doi: 10.1038/s41467-019-13185-3

Xu, P., Wu, L., Cao, M., Ma, C., Xiao, K., Li, Y., et al. (2021). Identification of MBW complex components implicated in the biosynthesis of flavonoids in woodland strawberry. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.774943

Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M., et al. (2021). MYB-mediated regulation of anthocyanin biosynthesis. *Int. J. Mol. Sci.* 22, 3103. doi: 10.3390/ijms22063103

Yao, G., Ming, M., Allan, A. C., Chao, G., Li, L., Xiao, W., et al. (2017). Map-based cloning of the pear gene *MYB114* identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J.* 92, 437–451. doi: 10.1111/tpj.13666

Yeh, K. C., Wu, S. H., Murphy, J. T., and Lagarias, J. C. (1997). A cyanobacterial phytochrome two-component light sensory system. *Science* 277, 1505–1508. doi: 10.1126/science.277.5331.1505

Yuan, F., Yang, H., Xue, Y., Kong, D, Ye, R, Li, C, et al. (2014). OSCA1 mediates osmotic-stress-evoked Ca2+ increases vital for osmosensing in arabidopsis. *Nature* 514 (7522), 367. doi: 10.1038/nature13593

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an r package for comparing biological themes among gene clusters. *Omics* 16, 284–287. doi: 10.1089/omi.2011.0118

Zoratti, L., Karppinen, K., Escobar, A. L., Häggman, H., and Jaakola, L. (2014). Light-controlled flavonoid biosynthesis in fruits. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00534

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 October 2022
DOI 10.3389/fpls.2022.1049114

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Shi Liu,
Northeast Agricultural University, China
Tongkun Liu,
Nanjing Agricultural University, China

*CORRESPONDENCE
Dexi Sun
sundexi@caas.cn
Junpu Liu
liujunpu@caas.cn

†These authors have contributed equally to this work

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

RECEIVED 20 September 2022
ACCEPTED 28 September 2022
PUBLISHED 19 October 2022

CITATION
Zhu Y, Yuan G, Wang Y, An G, Li W, Liu J and Sun D (2022) Mapping and functional verification of leaf yellowing genes in watermelon during whole growth period.
Front. Plant Sci. 13:1049114.
doi: 10.3389/fpls.2022.1049114

Mapping and functional verification of leaf yellowing genes in watermelon during whole growth period

Yingchun Zhu[1,2†], Gaopeng Yuan[1†], Yifan Wang[1], Guolin An[1], Weihua Li[1], Junpu Liu[1,2*] and Dexi Sun[1*]

[1]The Key Laboratory of Genetic Resource Evaluation and Application of Horticultural Crops (Fruit), Ministry of Agriculture, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China, [2]Western Research Institute, Chinese Academy of Agricultural Sciences, Changji, China

Increasing light energy utilization efficiency is an effective way to increase yield and improve quality of watermelon. Leaf is the main place for photosynthesis, and the color of leaf is directly related to the change of photosynthesis. In addition, leaf yellowing can be used as a marker trait to play an important role in watermelon hybrid breeding and improve seed breeding. It can not only be used to eliminate hybrids at seedling stage, but also be used to determine seed purity. In this study, transcriptome analysis was first carried out using the whole growth period leaf yellowing watermelon mutant *w-yl* and inbred line ZK, and identified 2,471 differentially expressed genes (DEGs) in the comparison group *w-yl*-vs-ZK. Among the top 20 terms of the gene ontology (GO) enrichment pathway, 17 terms were related to photosynthesis. KEGG pathway enrichment analysis showed that the most abundant pathway was photosynthesis—antenna proteins. The F_2 population was constructed by conventional hybridization with the inbred line ZK. Genetic analysis showed that leaf yellowing of the mutant was controlled by a single recessive gene. The leaf yellowing gene of watermelon located between Ind14,179,011 and InD16,396,362 on chromosome 2 by using indel-specific PCR markers, with a region of 2.217 Mb. In the interval, it was found that five genes may have gene fragment deletion in *w-yl*, among which *Cla97C02G036010*, *Cla97C02G036030*, *Cla97C02G036040*, *Cla97C02G036050* were the whole fragment loss, and *Cla97C02G0360* was the C-terminal partial base loss. Gene function verification results showed that *Cla97C02G036040*, *Cla97C02G036050* and *Cla97C02G036060* may be the key factors leading to yellowing of *w-yl* leaves.

KEYWORDS

Watermelon, leaf yellowing, gene mapping, fragment deletion, gene function

Introduction

Photosynthesis is essential in the process of plant growth and development, which is of great significance for plant survival. Leaves are the main place for photosynthesis in plants, and leaf color determines photosynthetic efficiency to a large extent (Chen et al., 2022). Different pigments can absorb light waves of different lengths, so leaves of plants show different colors due to different pigment contents and proportions. Leaf color mutation is a frequent and easily recognized phenomenon in nature, so leaf color mutants are ideal materials for studying plant development (Yuan et al., 2022b). At present, mutant materials have been found in a variety of plants, and the leaf color mutation types include albino, etiolation, stripe, yellow-green, green-yellow, green-white, light green and verdant green, etc (Awan et al., 1980). There are many ways of forming leaf color mutation. External factors mainly include light, temperature, plant hormones, mineral elements and metal ions. Internal factors mainly include genes related to photosynthetic pigment metabolism pathway, such as chloroplast biosynthesis pathway, chlorophyll degradation pathway, heme metabolism pathway and carotenoid metabolism pathway; as well as genes related to chloroplast development, such as chloroplast development and protein synthesis, nucleoplasmic interactions. All of these can lead to a decrease in the chlorophyll content of plant leaves, resulting in the leaves can not appear green color (Zhang et al., 2006; Sugliani et al., 2016; Li et al., 2018).

Studies on leaf color mutations mainly focus on the cell structure, photosynthetic physiology, molecular biology and other aspects of leaf color mutants, among which more in-depth studies have been conducted in model plants such as rice and *Arabidopsis*. For example, more than 160 leaf color mutants have been found in rice, distributed on 12 chromosomes, among which a small number of leaf color mutants have been cloned (Dong et al., 2013; Huang et al., 2017; Tan et al., 2019). Among them, 14 genes are directly involved in chlorophyll biosynthesis and catabolism (Sakuraba et al., 2013), and 6 genes are indirectly involved in this process (Yang et al., 2011), while 16 genes are directly involved in chloroplast development regulation (Gothandam et al., 2005) and 3 were indirectly involved in this process (Hibara et al., 2009) Therefore, the mutant genes are mainly divided into two categories, namely, genes in the chlorophyll biosynthesis and degradation pathway and genes in the chloroplast development pathway. In addition, previous studies have proved that most leaf color mutations are nuclear inheritance except for a small number of leaf color mutations for cytoplasmic inheritance (Kong et al., 2016; Li et al., 2021a; Li et al., 2021b). In recent years, with the application of high-throughput sequencing, the study of leaf color mutation has been gradually carried out in some important economic crops and ornamental plants, such as tea, pepper, maize, melon and cucumber (Shao et al., 2013; Li, 2016; Lai et al., 2018; Wang et al., 2019; Zhu et al., 2019; Gao et al., 2020; Xiong et al., 2020), which will help improve crop quality and increase yield (Shao, 2013; Ren et al., 2019). The results of the latest study on cucumber showed that the post-green mutant *SC311Y* was controlled by a recessive gene, which was identified as the gene controlling chloroplast development by BSA-seq and RNA-seq techniques (Zhang et al., 2022).

The genetic basis of watermelon is narrow and the natural mutation rate is low. There are few studies on watermelon leaf color mutants. The leaf color mutation materials are mainly divided into four categories: (1) watermelon leaf color mottled mutants, which are characterized by white-green cotyledons and mosaic-like spots in the first true leaf under low temperature environment (Provvidenti, 1994; Wang et al., 2011); (2) watermelon albino mutant, showing pale yellow or pale cream cotyledons, gradually turning green but remaining white at leaf margins, white tendrils, petioles, petals and hypocotyls (Zhang et al., 1996; Wang et al., 2011; Hou et al., 2016); (3) incomplete dominant yellow leaf mutants (Hou et al., 2016); (4) In post-green mutants, the leaves showed light green cotyledons and leaves at the early stage, and changed to normal green at the later stage (Wang and Wang, 1997; Ma and Zhang, 1999; Wang et al., 2011; Xu et al., 2022). In terms of genetic analysis and molecular biology, the early stage mainly focused on the study of genetic patterns, and confirmed that watermelon leaf color mutants were controlled by recessive genes based on the discovered mutant materials (Rhodes, 1986; Provvidenti, 1994; Zhang et al., 1996). With the publication of watermelon genome and the rapid development of sequencing technology (Guo et al., 2019; Wu et al., 2019), more high density genetic maps of watermelon emerged (Duan et al., 2022), but only a few maps involved watermelon leaf color. For example, Haileslassie (Haileslassie, 2020) found the presence of a SNP in the gene *ClCG03G010030* of the watermelon post-green mutant *Houlv*, resulting in an arginine to lysine mutation. The gene encodes an FtsH extracellular protease family protein which is involved in the development of early chloroplast. Exploring the mechanism of leaf color variation can provide a theoretical basis for genetic improvement and meet people's needs in production, seed selection and breeding.

China is the largest watermelon planting and consumption country in the world. Although the demands for watermelon is diversified, cultivating new varieties with high yield and high quality is still the main direction of watermelon breeding. Improving the utilization efficiency of light energy of watermelons is an effective way to promote yield and improve quality. In this study, yellow leaf throughout the whole growth period material *w-yl* and green leaf material ZK were used as experimental materials. The position of the leaf yellowing gene in the chromosome was preliminarily located by BSA-seq technology. The high-density genetic map was constructed by the F_2 population using InDel markers for mapping the position of the mutant gene in the chromosome, and the key candidate

genes and key variations were screened in combination with transcriptome data. Finally, the virus-induced gene silencing (VIGS) assay was performed on the key candidate genes to clarify the function of the yellowing leaf gene. The development of this study will help to explore the mechanism of leaf yellowing in the whole growth period of watermelon, and provide theoretical support for the application of leaf yellowing and molecular marker-assisted selection of new watermelon varieties with high photosynthetic efficiency.

Materials and methods

Plant material cultivation and samples collection

The leaf color yellowing mutant material *w-yl* was obtained from the National Mid-term Genebank for Watermelon and Melon (Zhengzhou, China), the leaves in the whole growth period were yellow, including cotyledon and fruit. Normal green leaf material ZK was supplied by the Diploid Watermelon Genetics and Breeding Research Group of Zhengzhou Fruit Research Institute (ZZFRI) of Chinese Academy of Agricultural Sciences (CAAS). In this study, the mutant material was crossed with the ZK, and six generations were constructed: P_1 (the yellow parent *w-yl*), P_2 (the green leaf parent ZK), F_1 (orthogonal), BC_1P_1, BC_1P_2, and F_2. The materials were planted in a greenhouse at the Xinxiang Comprehensive Experimental Base of CAAS, with a row spacing of 1.5 m and a plant spacing of 0.4 m. The phenotype of leaf color was determined by visual observation.

Plant for chlorophyll were planted in an artificial climate chamber and treated with different environmental factors at three true-leaf stage: temperature 35°C/28°C, light intensity 30,000 Lx, namely HTHL(high temperature and high light); temperature 35°C/28°C, light intensity 12,000 Lx, namely as HTNL (high temperature and normal light); temperature 35°C/28°C, light intensity 5,000 Lx, marked as HTLL (high temperature and low light); temperature 28°C/25°C, light intensity 30,000 Lx, marked as NTHL (normal temperature and high light); temperature 28°C/25°C, light intensity 12,000 Lx, marked as NTNL (normal temperature and normal light); temperature 28°C/25°C, light intensity 5,000 Lx, marked as NTLL (normal temperature and low light); temperature 15°C/15°C, light intensity 30,000 Lx, marked as LTHL (low temperature and high light); temperature 15°C/15°C, light intensity 12,000 Lx, marked as LTNL (low temperature and normal light); 15°C/15°C, 5,000 Lx, labeled as LTLL (low temperature and low light). Light cycle was 16h/8h, humidity 80%. Each treatment set three replicates. Chlorophyll content was determined after 8 days of treatment.

Plant for chlorophyll precursors and transcriptome sequencing were grown in a smart greenhouse in ZZFRI of CAAS, the light cycle was 16h/8h, the temperature was 25°C/18°C and the light is natural light. The leaves were sampled after 8 days of treatment.

Determination of pigment content

The third true leaf from five seedlings was sampled and mixed, weighed 0.1 g and put into a 15 mL centrifuge tube respectively, added 10 mL of 96% ethanol, and soaked in dark environment until the leaves turned completely white (Yuan et al., 2017). The absorbance A665, A649 and A470 at 665 nm, 649 nm and 470 nm were determined by UV spectrophotometer (UV-2600I, Shimadzu, Kyoto, Japan). The concentrations of chlorophyll a (chla), chlorophyll b (chlb), total chlorophyll (chla+b) and carotenoids were calculated using 96% ethanol as blank control. The equations are as following:

$$Chla\left(mg\cdot L^{-1}\right) = 13.95 \times A_{665} - 6.88 \times A_{649}$$

$$Chlb\left(mg\cdot L^{-1}\right) = 24.96 \times A_{649} - 7.32 \times A_{665}$$

$$Chla+b\left(mg\cdot L^{-1}\right) = chla + chlb = 6.63 \times A_{665} + 18.08 \times A_{649}$$

$$Carotenoids\left(mg\cdot L^{-1}\right) = \left(1000 \times A_{470} - 2.05 \times Chla - 114.8 \times Chlb\right)/248$$

Chlorophyll content ($mg\cdot g^{-1}$ = (C × V)/(W × 1000). C represents chlorophyll content, V represents the total volume of extract (mL), and W represents leaf mass (g).

Determination of chlorophyll precursor

The contents of main chlorophyll precursor in the process of chlorophyll synthesis were measured, among which δ-aminolevulinic acid (ALA) was determined according to the method of Dei (Dei, 2010) and the molar concentration of ALA was calculated with a molar extinction coefficient of 7.2×10^4 $mol^{-1}\cdot cm^{-1}$at 535 nm. Relative contents of protoporphyrin IX (protoIX), Mg-protoporphyrin IX (Mg-proto IX), and pchlide were determined according to the method of Rebeiz (Rebeizjames et al., 1975) and Lee (Lee et al., 1992). The relative mass molar concentration of Mg-Proto IX is presented as $F_{\frac{440}{ex595}}$: fluorescence emission intensity at 595 nm under 440 nm excitation light. Proto IX $(F_{\frac{440}{ex633}})$ = $(F_{\frac{440}{ex633}} - 0.25 \times F_{\frac{440}{ex622}} - 0.24 \times F_{\frac{440}{ex640}})/0.95$;
$Pchlide\left(\frac{440}{ex640}\right) = \left(F_{\frac{440}{ex640}} - 0.03 \times F_{\frac{440}{ex633}}\right)/0.99$.

RNA sequencing

Leaves of five plants were selected as a sample from *w-yl* and ZK, with three biological replicates respectively. Total RNA was extracted using RNeasy Plant Mini Kit (Beijing Tiagen), following the manufacturer's instructions. Then RNA was reversely transcribed to cDNA, and the cDNA fragments were segmented by PCR. Finally, the double-stranded PCR product is thermally denatured to form single-stranded circular DNA, which is then formatted into a final library. The cDNA library was sequenced by BGISEQ-500 system (BGI-Shenzhen, China) with reads of 100bp in length.

The sequencing data were screened to obtain Clean reads, which were then mapped into the '97103' watermelon genome (http://cucurbitgenomics.org/organism/21) using Bowtie2. Gene expression levels were calculated using FPKM (million fragments per kilobase). Based on KEGG (http://www.genome.jp/kegg/) and GO (http://www.geneontology.org/) database for gene annotation and function assignment. Differentially expressed genes (DEGs) were set as gene fold change ≥2.00 and false discovery rate ≤0.001. Through GO enrichment and KEGG enrichment pathways, the significantly enriched metabolic pathways were screened and compared with the whole genome background. Functional classification of DEGs was performed according to GO and KEGG annotation results and official classification, and FDR ≤ 0.01 was set as significant enrichment.

QPCR validation and gene expression analysis

Total RNA was extracted by plant RNA kit (Huayue Yang Biotechnology Co., LTD.). A total of 1.0 μg of RNA was used for cDNA synthesis using the PrimeScript RT kit and gDNA Eraser (TaKaRa) according to the manufacturer's protocol. Primers were designed using NCBI online tools (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), and synthesized by Sangon Biotech (Shanghai, China). All the primer sequences were shown in Supplementary Table 1. Quantitative real-time PCR reaction procedure and system were as described previously (Yuan et al., 2022a). All primers are shown in Supplementary Table S1. The $2^{-\Delta\Delta Ct}$ method was used to calculate relative gene expression values (Kenneth and Thomas, 2002).

BSA-seq analysis of the leaf yellowing genes

Leaf DNA of 30 individual plants with yellowed and green extreme phenotypes in F2 population were selected for the construction of two extreme sequencing mixed pools, and parental DNA was used to construct the parental pools for sequencing analysis. The depth of parental sequencing was 20×, and the depth of extremely mixed-pool sequencing was 30×. Sequencing was performed by Biomarker Technologies Co, LTD (Beijing, China) using Illumina HiSeq2000. The sequencing read length was 150 bp.

Raw reads were filtered to remove reads containing adapter, and reads containing >5% N and low-quality reads (the number of bases with quality value Q ≤ 10 accounted for more than 50% of the whole read) were used to obtain clean reads for subsequent analysis. Clean reads were mapped to the '97103' watermelon genome (http://cucurbitgenomics.org/organism/21) using BWA software. Then GATK (4.0.4.0) and SNPeff (4.3) were used to annotate the mutation sites, and single nucleotide polymorphisms (SNPs) and insertion-deletion polymorphisms (InDels) were identified.

The SNP-index algorithm was used to establish the target region to find the significant difference in genotype frequency between the pool, and Δ(SNP-index) was used for statistics. In this project, the DISTANCE method was used to fit the ΔSNP-index, and then the region above the threshold was selected as the region related to the trait according to the association threshold. The stronger association between SNP and trait, the closer Δ(SNP-index) to 1.

Functional analysis of key genes

Using the cDNA of green leaf ZK as template, specific primers were designed to amplify the CDS regions of *Cla97C02G036010*, *Cla97C02G036030*, *Cla97C02G036040*, *Cla97C02G036050* and *Cla97C02G036060*, and the primers were shown in Supplementary Table 1. BamHI (GGATCC) restriction sites were added to both ends of the primers and inserted into the cucumber green mottle mosaic virus (CGMMV) gene silencing vector PV190 by homologous recombination to construct virus-induced gene silencing (VIGS) vector. The dual vector was transformed into *Agrobacterium tumefaciens* GV3101.

Induction and inoculation of *A. tumefaciens* according to Liu (Liu, 2019) When watermelon seedlings were at cotyledon stage, the induced *A. tumefaciens* was injected from the back of watermelon cotyledon with 1 mL syringe. The blank control (Blank, B), water control (Water, W), medium control (YT medium, Y), blank vector control (PV190, P) and *PDS* gene positive control (PDS) were set up respectively. Three biological replicates were set up for each treatment. Two weeks after injection, leaf phenotype was observed, and samples were collected for ultrastructural analysis, chlorophyll content measurement and gene expression analysis.

Ultrastructural observation of chloroplast

The above-mentioned leaves with phenotype after *A. tumefaciens* were used as materials, fixed with 4% glutaraldehyde (configured with pH 7.2 phosphate buffer) overnight at 4°C, rinsed with phosphate buffer three times, fixed with 1% osmium tetroxide for 1 h, rinsed with phosphate buffer three times, dehydrated with 30%, 50%, 70%, 80%, 95%, 100% ethanol and acetone step by step for 5 min, and finally embedded with resin. After sectioning, they were stained with 2% uranyl acetate saturated alcohol and lead citrate for 15 min, and the chloroplast ultrastructure was observed under transmission electron microscope (HT7700, Hitachi, Japan).

Data statistical analysis

All data graphs were analyzed by Office 2016 software. Differences were analyzed by SPSS 18.0 software, and one-way ANOVA was used for statistical analysis, $p< 0.05$ ($n = 3$) was considered significant difference.

Results

Genetic characteristics analysis of yellowing leaf color

The leaves of *w-yl* showed yellow throughout the whole growth period (Figure 1A), and the color did not change with environmental changes, such as temperature and light intensity (Figure 1B). Under different temperature and light intensity, there were no significant differences in the contents of chla, chlb, chla+b and carotenoids.

In addition, phenotypic data showed that all F_1 plants appeared green leaves, indicating that the yellow mutation was recessive. For F_2 plants, among the 237 progeny in the summer of 2018, 178 plants had green leaves and 59 plants had yellow leaves; among the 993 progeny in spring of 2019, 730 had green leaves and 263 had yellow leaves (Table 1). The χ^2 test of green and yellow leaves in the two seasons showed that the separation pattern was consistent with the Mendelian separation ratio of 3:1 ($\chi^2 > \chi^2_{0.05} = 3.841$). Furthermore, for the backcross progeny BC_1P_1 and BC_1P_2, the yellow leaf plants were 29 and 0 respectively, indicating that the yellowing mutation of watermelon leaves conformed to the genetic pattern controlled by a single recessive nuclear gene, and green leaves were dominant to yellowing.

Genetic characteristics analysis of yellowing leaf color

Previous studies had demonstrated that there are significant differences in chlorophyll content and photosynthetic indicators (Ren et al., 2019). To further validate the difference, the chlorophyll precursors, including ALA, protoIX, Mg-ProtoIX and pchlide were analyzed (Figure 2). The results showed that the contents of four indexes detected in the *w-yl* were significantly lower than those in ZK, which explained the low chlorophyll content to a certain extent.

RNA-seq for the leaves of *w-yl* and ZK

A total of 6 samples were measured by RNA-seq, including 3 samples for *w-yl* and 3 samples for ZK, yielding an average of 6.06 Gb of data per sample. The average rate of genome alignment was 89.40%, and the average rate of gene set alignment was 65.81% (Figure 3A). For the comparison group *w-yl*-vs-ZK, a total of 19,261 genes were detected, and there were 18,323 shared genes, including 2,471 DEGs (Figure 3A), with 848 up-regulated DEGs and 1893 down-regulated DEGs (Figure 3B).

GO enrichment and KEGG pathway enrichment analyses were carried out to better understand the function of DEGs. For GO enrichment, 17 of the top 20 selected GO terms were related to photosynthesis process, including 4 terms related to photosystem, such as photosystem (GO:0009521), photosystem I (GO:0009522), Photosystem II (GO:0009523), light harvesting in photosystem I (GO:0009768); 4 terms involved in photosynthesis, such as photosynthesis (GO:0015979), photosynthetic membrane (GO:0034357), photosynthesis—light reaction (GO:0019684) and photosynthesis—light harvesting (GO:0009765); 6 terms involved in thylakoid, such as thylakoid (GO:0009579), thylakoid membrane (GO:0042651), chloroplast thylakoid (GO:0009534), chloroplast thylakoid membrane (GO:0009535), plastid thylakoid (GO:0031976) and plastid thylakoid membrane (GO:0055035); 3 terms relate to pigments, such as tetrapyrrole binding (GO:0046906), chlorophyll binding (GO:0016168) and protein-chromophore linkage (GO:0018298) (Figure 3C; Supplementary Table S2). These results showed that *w-yl* and ZK had significant differences in photosynthesis.

For KEGG pathway enrichment, the two most significant enrichment pathways of the top 20 pathways were photosynthesis—antenna proteins and plant—pathogen interaction (Figure 3D; Supplementary Table S3). Due to the importance of antenna protein for photosynthesis (Figure 4A), we focused on the analysis of antenna protein-related DEGs, and completely screened 16 DEGs that encoded antenna protein (Figure 4B). LHCI and LHCII, as important components of photosystem I complex and photosystem II complex, are composed of four and six small components, respectively. For LHCI, the number of DEGs that encoded LHCI Chl a/b binding protein 1 (Lhca1), Lhca2, Lhca3 and Lhca4 was 1, 2, 1 and 2, respectively. For LHCII, the number of DEGs that encoded LHCII Chl a/b binding protein 1 (Lhcb1), Lhcb2, Lhcb3, Lhcb4, Lhcb5 and Lhcb6 was 4, 1, 1, 2, 1 and 1, respectively. The expression levels of all the 16 DEGs in ZK were significantly higher than those of *w-yl*, and the fold change was between 2.6 and 14.0 (Figure 4B).

10.3389/fpls.2022.1049114

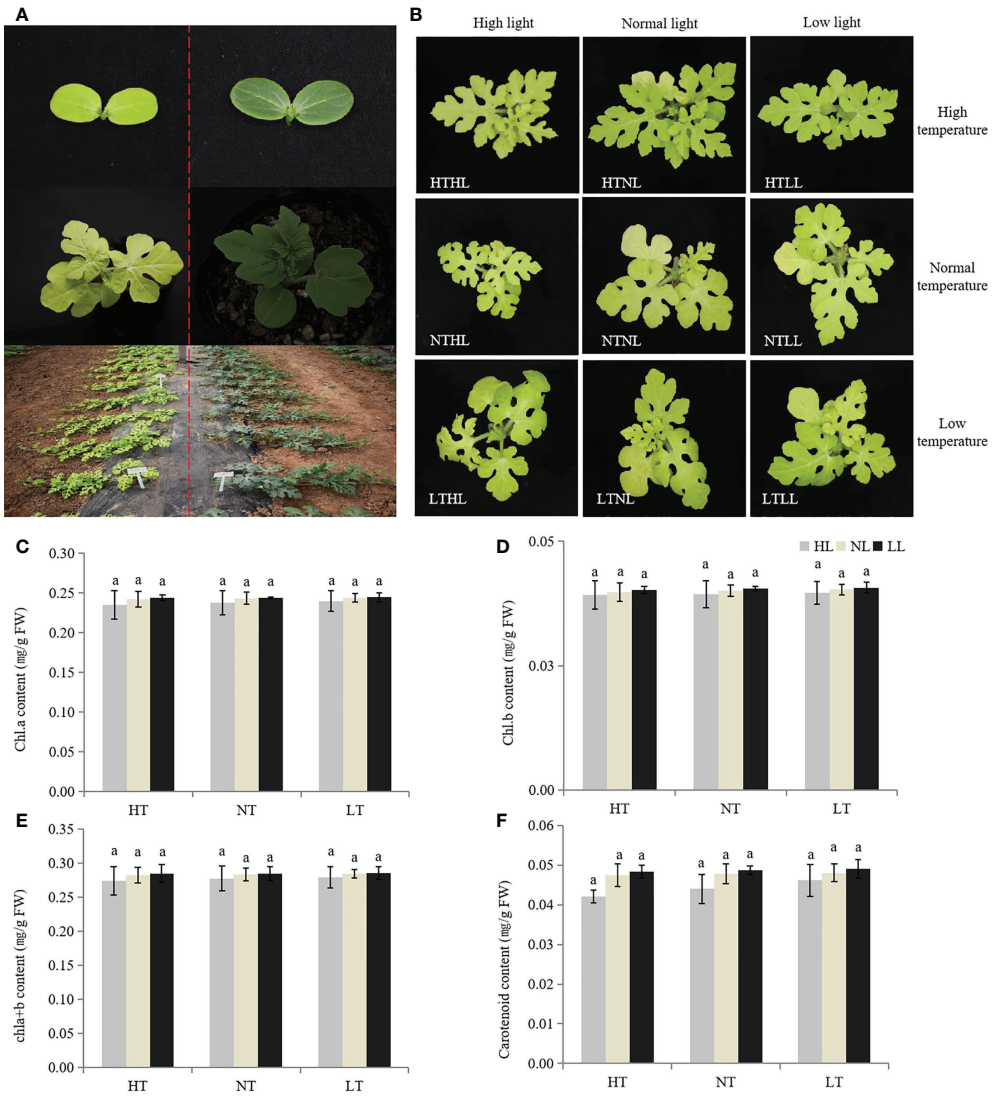


FIGURE 1
Plant phenotypes of w-yl and ZK. **(A)** Plant phenotypes of w-yl (left) and ZK (right) at different developmental stages. **(B)** Plant phenotypes of w-yl under different temperature and light intensity. The content of **(C)** chla, **(D)** chl b, **(E)** chla+b and **(F)** carotenoid under different temperature and light intensity. Small letters represent significant difference at P<0.05.

TABLE 1 Phenotype of yellow mutant to green leaf trait and Chi-square goodness-fit test ratios in different populations.

Population	Number	Green leaves	Yellow leaves	Expected ratio		P value
P_1	15		15			
P_2	15	15				
F_1	30	30				
F_2 (Summer of 2018)	237	178	59	3:1	0.0014	0.9701
F_2 (Spring of 2019)	993	730	263	3:1	1.1685	0.2797
BC_1P_1	54	29	25	1:1	0.2963	0.5862
BC_1P_2	30	30	0			

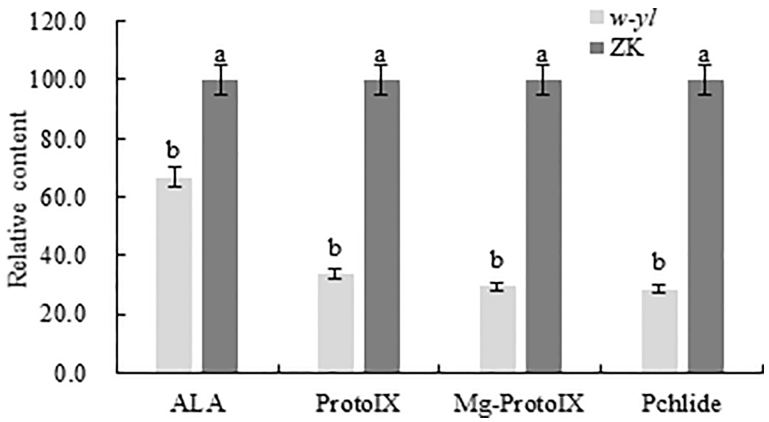


FIGURE 2
Chlorophyll precursors content of the *w-yl* and CK in pepper leaf. Small letters represent significant difference at P<0.05.

Verification of DEGs of qPCR and RNA-seq data

To verify the accuracy of RNA-seq data, 12 DEGs (*Cla97C02G035950*, *Cla97C02G035960*, *Cla97C02G035980*, *Cla97C02G036070*, *Cla97C02G036090*, *Cla97C02G036110*, *Cla97C02G036130*, *Cla97C02G036140*, *Cla97C02G036150*, *Cla97C02G036160*, *Cla97C02G036190* and *Cla97C02G036200*) of the 29 genes in the interval were selected to conduct qPCR (Figure 5). The results showed that expression patterns of 12 DEGs were highly consistent with those of genes in RNA-seq data, which demonstrated that the RNA-seq data are reliable.

Mapping of yellowing gene in w-yl leaf

In order to quickly identify the key candidate genes related to leaf color in the F_2 population, 30 green and 30 yellow leaf

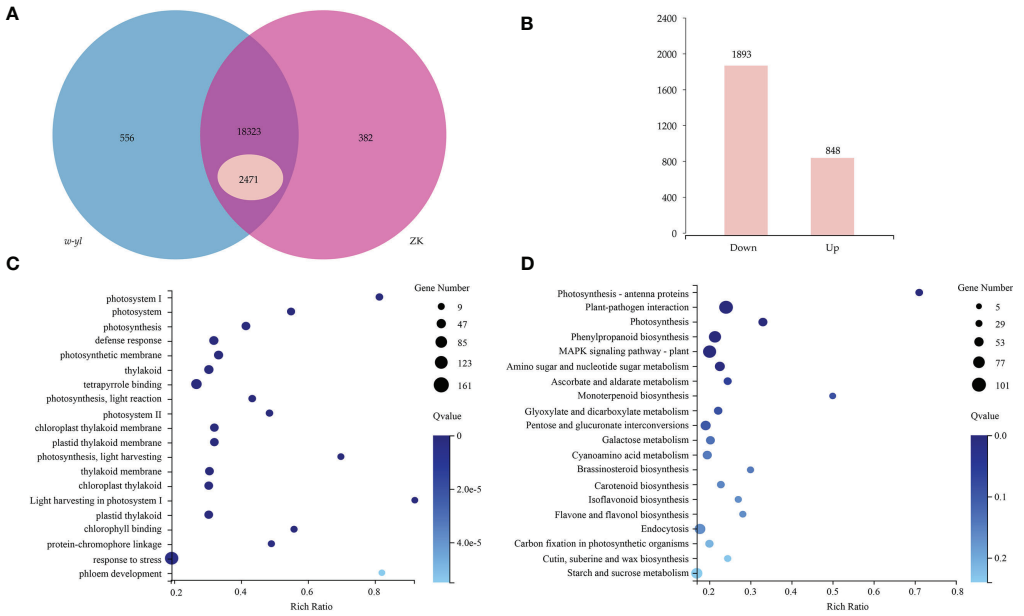


FIGURE 3
RNA-seq analysis for the leaves of *w-yl* and ZK. **(A)** Venn diagram of the relationship of *w-yl*-vs-ZK. **(B)** Number of up-regulated and down-regulated DEGs. **(C)** GO enrichment analysis of DEGs. **(D)** KEGG pathenrichment analysis of DEGs.

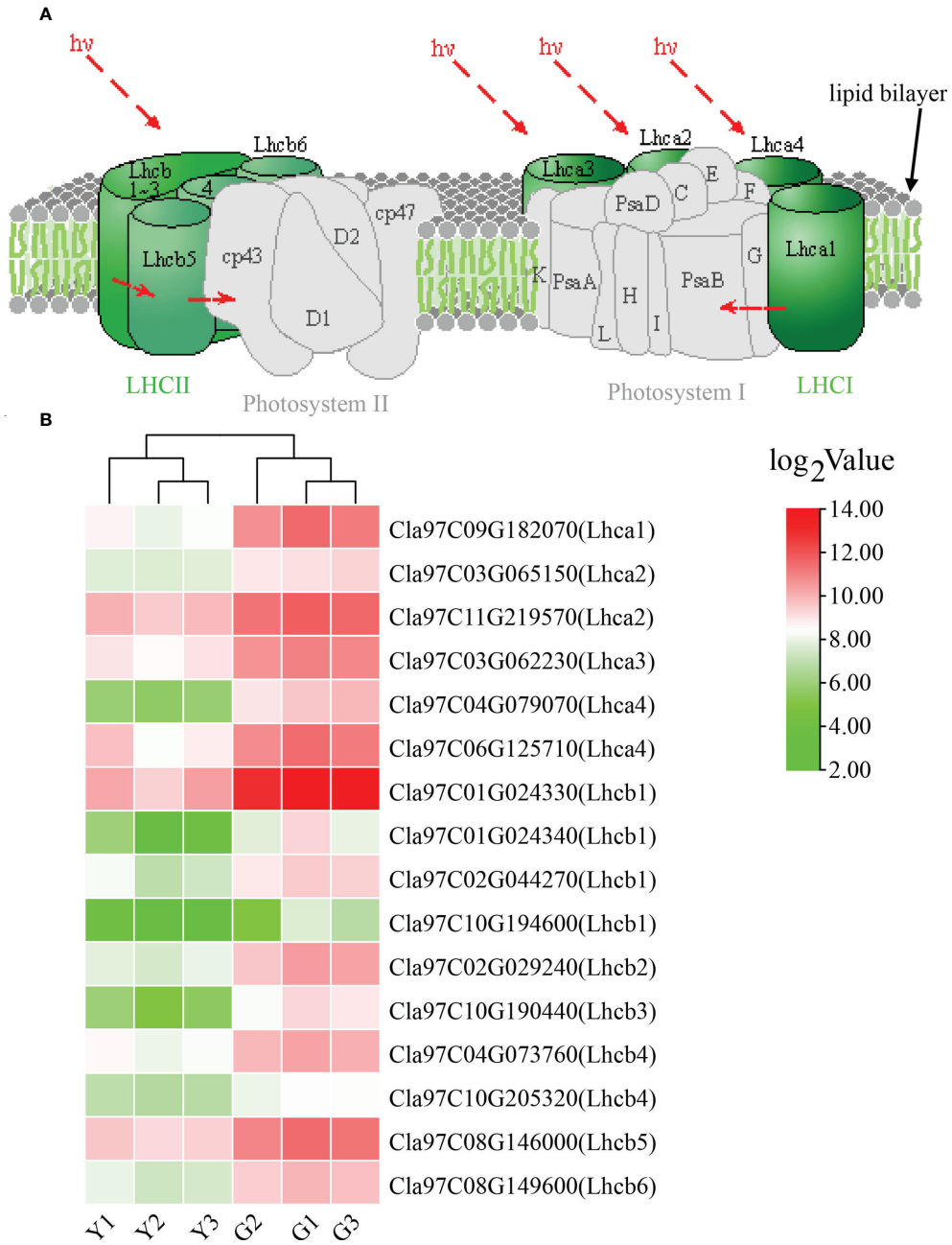


FIGURE 4
Analysis of the photosynthesis—antenna proteins. **(A)** Role of photosynthesis—antenna proteins in photosynthetic system. **(B)** Expression of photosynthesis—antenna proteins related DEGs.

progeny were selected and sequenced on the Illumina platform. A total of 51.0 Gb clean bases were generated with an average depth of about 26.5×. Finally, we identified 266,255 SNPs between *w-yl* and ZK, and 83,373 SNPs between the F_2 pools. According to the SNP-index values of *w-yl* and ZK, the Δ(SNP-index) value of approximately 7.42 Mb genome region (11,540,000-18,960,000) on chromosome 2 was greater than the threshold (Figure 6A). These results indicated that this region might contain the key gene of watermelon leaf yellow traits.

In order to further locate the candidate genes for yellowing leaf, the chromosome region of the variation between *w-yl* and ZK were analyzed. A total of 12 pairs of InDel molecular markers (Supplementary Table S4) were developed for the candidate region

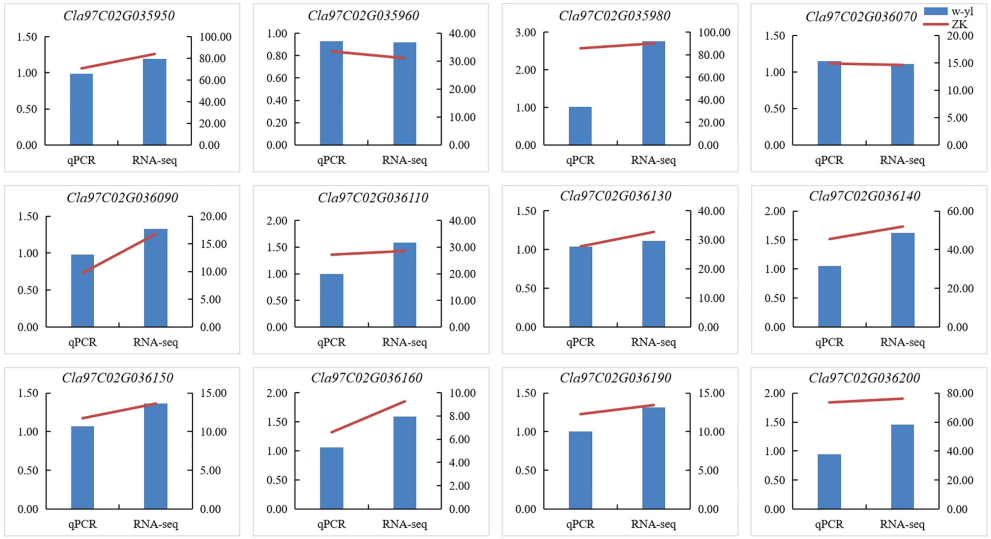


FIGURE 5
Verification of DEGs by RT-qPCR.

based on 233 F_2 populations. The results verified that these genes were in the range of 11.54 Mb—18.96 Mb on chromosome 2. Subsequently, based on the determination of leaf color phenotype data and individual exchange genotype, 12 recombinant individuals were further screened using 1883 F_2 populations. Finally, it is found that the candidate interval corresponds to the 2.217 Mb region of InD14,179,011—InD16,396,362 (Figure 6B). There were 29 genes in this region and annotated them according to the watermelon reference genome (Figure 3C; Table 2). Notably, compared with ZK, *Cla97C02G036010*, *Cla97C02G036020*, *Cla97C02G036030*, *Cla97C02G036040*, and *Cla97C02G036050* were completely absent in *w-yl*, and *Cla97C02G036060* had partial base deletion, suggesting that they were the key genes determining *w-yl* leaf color mutation (Supplementary Figure S1).

In addition, the results of agarose gel electrophoresis and qPCR showed that *Cla97C02G036010, Cla97C02G036020, Cla97C02G036030, Cla97C02G036040, Cla97C02G036050* and *Cla97C02G036060* could not be amplified in *w-yl*, as well as *Cla97C02G036020* also had no target product in ZK (Figure 7). The RNA-seq results also showed the same results (Figure 6C). These results further proved the importance of *Cla97C02G036010, Cla97C02G036030, Cla97C02G036040, Cla97C02G036050* and *Cla97C02G036060* in leaf yellowing.

Function analysis of yellowing gene in ZK leaf

In order to verify the gene function of the candidate genes, cucumber mosaic virus-mediated VIGS vector was used to perform gene silencing assay on ZK leaves. The results showed that at 16 days after inoculation (DAI), the plants inoculated with water (Figure 8B), medium (Figure 8C) and blank vector (Figure 8D) showed no significant difference in phenotype compared with the blank control (Figure 8A), while the positive control plants inoculated with *PDS* gene showed virus symptoms at DAI16, with severe true leaf pucking and chlorosis (Figure 8E). Watermelon plants silencing *Cla97C02G036010* (Figure 8F) and *Cla97C02G036030* (Figure 8G) showed symptoms of disease at DAI17, and their true leaves were slightly wrinkled and mottled greenish yellow. Watermelon plants silencing *Cla97C02G036040* (Figure 8H), *Cla97C02G036050* (Figure 8I) and *Cla97C02G036060* (Figure 8J) showed obvious virus symptoms at DAI13, with obvious true leaf wrinkling and large area mottled yellow.

Then the expression levels of the silenced genes were detected, when compared with the control group (B, W, Y, P and PDS), their expression levels were significantly reduced. Among them, the expression of *Cla97C02G030640* decreased most sharply, which were 2.9%, 2.9%, 2.8%, 3.2% and 27% of the control group, respectively (Figure 9). Besides, the results of chlorophyll content of leaves with phenotype showed that there was no significant difference in chla, chlb and chla+b content among groups B, W, Y and P, while the chlorophyll contents of silenced *PDS* group was significantly lower than that of the former four groups. For the five silenced genes, the contents of chla, chlb and chla+b were significantly lower than those of B, W, Y and P control groups. Among them, the contents of chla, chlb and chla+b in silenced genes *Cla97C02G030640* and *Cla97C02G030660* were the lowest and had not significantly different from those in silenced *PDS* group (Figure 10).

Furthermore, the ultrastructure of chloroplast was analyzed to analyze the reasons for these phenomenon. As a result, the

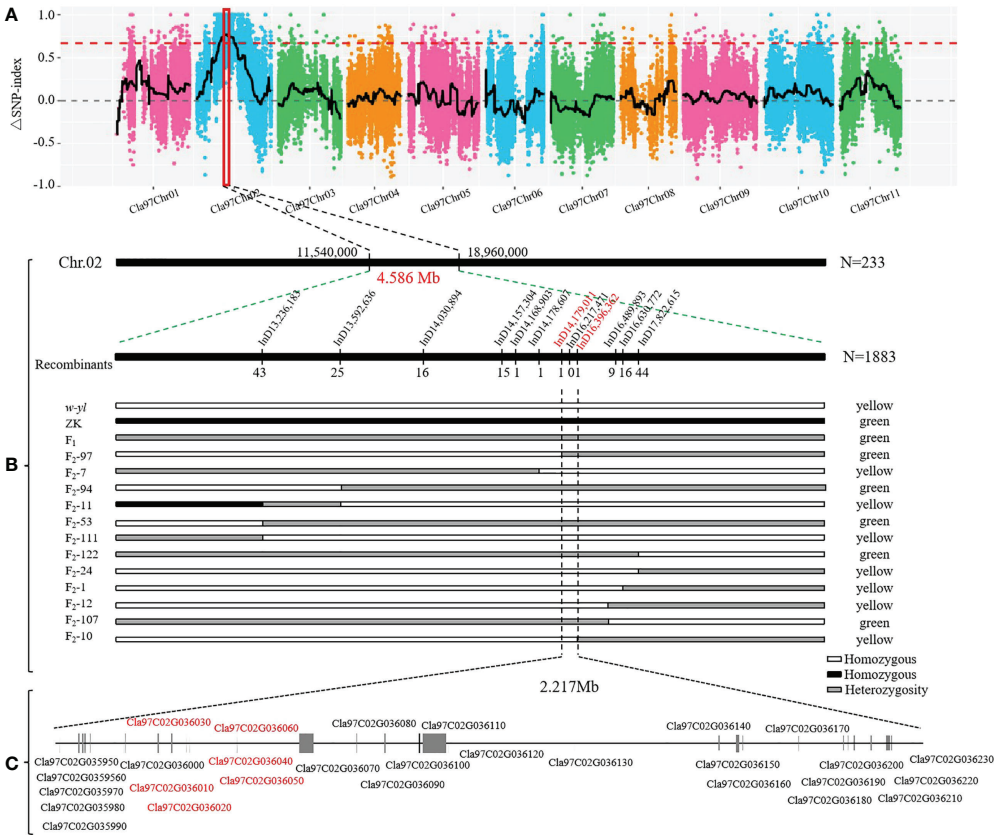


FIGURE 6
Location of yellowing gene on watermelon chromosome 2. **(A)** Δ(SNP-index) of watermelon chromosomes. **(B)** The candidate genes was mapped to a 4.586 Mb region between InD14,179,011 and InD16,396,362 on chromosome 2. **(C)** Putative genes in the candidate region based on the watermelon reference genome annotation.

chloroplast ultrastructure of the silenced gene *Cla97C02G030610* (Figure 11B), *Cla97C02G030630* (Figure 11C) and *Cla97C02G030650* (Figure 11D) did not change compared with the blank control (Figure 11A), and all contained normal grana lamella (GL) and plastid globule (PL). However, the ultrastructure of silenced gene *Cla97C02G030640* and *Cla97C02G03060* was significantly changed, and the chloroplast structure may be damaged (red dotted circle area). For silenced gene *Cla97C02G030660* (Figure 11F), there was no PL, and PG stratification was not obvious, appearing in a fuzzy state. Especially for the silenced gene *Cla97C02G030640* (Figure 11D), there was no PL and no obvious PG, speculating that PG was degraded.

Taken together, these results indicated that candidate genes play an important role in causing leaf yellowing, especially *Cla97C02G030640*.

Discussion

There are various types of leaf color mutations, and leaf yellowing was the most common phenomenon (Jin et al., 2021). Plant leaf yellowing mutants, also known as chlorophyll deficiency mutants, are usually caused by the destruction of chlorophyll synthesis or degradation pathways (Yang et al., 2014). At present, yellowing mutants have been found in rice (Zhang et al., 2017), tomato (Yao et al., 2010) and rape (Xiao et al., 2013). In this study, we reported a whole growth period leaf yellowing watermelon material *w-yl* (Figure 1A), which is completely different from the published watermelon leaf color mutant material (Wang et al., 2011; Haileslassie, 2020), and the leaf yellowing characteristics of *w-yl* can be stably inherited. Light can affect plant chloroplast development and chlorophyll metabolism. For example, light intensity is very important for chloroplast formation (Franck et al., 2000), which can change the proportion and content of anthocyanins or chlorophyll or carotenoids by affecting the activity of enzymes related to pigment synthesis or the expression of genes related to photosynthesis, thus causing the color change of leaves, and eventually leading to the formation of leaf color mutants (Xu et al., 2021). In cucumber, the pigment content of the post-green mutant *SC311Y* increased significantly under lower light

TABLE 2 Gene function annotation information in candidate interval.

Gene name	Gene function
Cla97C02G035950	Translator-related TMA7
Cla97C02G035960	BZIP transcription factor, putative (DUF1664)
Cla97C02G035970	lipid-binding serum glycoprotein
Cla97C02G035980	Protein nucleo-fusion transmitter 6, chloroplast/ mitochondria-like isoform X1
Cla97C02G035990	Unknown protein
Cla97C02G036000	L-ascorbate oxidase homolog, Oxidoreductase activity, Cu^{2+} binding
Cla97C02G036010	Unknown protein
Cla97C02G036020	Two component response regulator like protein
Cla97C02G036030	Transmembrane protein, putative
Cla97C02G036040	Protein containing DUF679 domain
Cla97C02G036050	DnaJ homologous subfamily B member 13 like
Cla97C02G036060	Protein Ycf2
Cla97C02G036070	U11/U12 small ribonucleoprotein 65 kDa protein isoform X2
Cla97C02G036080	Unknown protein
Cla97C02G036090	RING-type E3 ubiquitin transferase
Cla97C02G036100	family proteins containing pentapeptide repeats
Cla97C02G036110	Niemann-Pick C1 protein-like isoform X2
Cla97C02G036120	Zinc finger family protein
Cla97C02G036130	Integral hemolysin III-like protein
Cla97C02G036140	Ser/Thr-rich T10 in the DGCR region
Cla97C02G036150	Phosphoglycerate mutagenase family proteins
Cla97C02G036160	SEC1 family transporter SLY1, Oxidoreductase activity, Mg^{2+} binding
Cla97C02G036170	Unknown protein
Cla97C02G036180	Retrotransposon protein, unclassified
Cla97C02G036190	Glycine-rich RNA-binding protein, putative
Cla97C02G036200	Plant UBX domain protein 4
Cla97C02G036210	Calcium-permeable stress-gated cation channel 1
Cla97C02G036220	Acid phosphatase/vanadium-dependent haloperoxidase-related protein
Cla97C02G036230	Core-2/I branch β-1,6-N-acetylglucosamine aminotransferase family proteins

conditions and was vulnerable to light (Zhang et al., 2022). In addition, the synthesis process of chlorophyll is regulated by many enzymes, and its activity is regulated by temperature (Yang et al., 2018). In rice, mutant *tcd9* showed abnormal chloroplasts and fewer thylakoid lamellae in albino mutant seedlings at low temperature, but the mutant showed normal green color at high temperature (Jiang et al., 2014). In *Arabidopsis*, a heat-sensitive mutant in *tsl1* is impaired in chloroplast RNA editing at high temperatures, hampering chloroplast development (Sun et al., 2020). However, under high temperature and low temperature, high light intensity and low light intensity, chlorophyll content and carotenoid content of *w-yl* had no significant difference compared with normal temperature and light intensity (Figures 1B–F), which suggesting that the mutant *w-yl* was non-photosensitive and non-temperature sensitive.

Previous study had confirmed that the chloroplast volume, the number of thylakoids and the number of grana lamellae in the leaves of mutant *w-yl* are smaller, which leads to a significant reduction in chlorophyll content (Ren et al., 2019). In fact, leaf yellowing mutations are usually caused by incomplete chloroplast development. For example, the yellow green leaf mutant *ygl8* in rice was caused by chloroplast dysplasia (Kong et al., 2016) The mutation of *ChlI/Chl9*^{pyl3} gene in rice leaded to the formation of *pyl3* mutant with light yellow leaves. which inhibited chlorophyll synthesis, resulting in chloroplast dysplasia and leaf color variation (Hu et al., 2021). In *Brassica napus*, the chloroplast morphology of the leaf yellowing mutant *S28-y* was abnormal, with no complete grana and grana lamellae, resulting in total chlorophyll deficiency (Ge et al., 2022). A large number of studies have shown that chlorophyll is the main factor affecting leaf color phenotype, and leaf color phenotype is closely related to chlorophyll content, and the proportion of photosynthetic pigments in leaves can be directly expressed by the depth of leaf color (Chen et al., 2017; Chen et al., 2018; Su et al., 2020). Chlorophyll precursor material is the intermediate product of chlorophyll synthesis process, any step of which will influence the chlorophyll content (Wang et al., 2009) (Beale, 2005). For example, in rice leaf yellowing mutant *W1*, the process from porphobilinogen to uroporphyrinogen III was blocked, which hindered the synthesis of chlorophyll (Cui et al., 2001). In addition, Kong et al. found that the *YGL8* gene isolated and identified in *ygl8* rice yellow-green leaf mutant can encode Mg-protoIX, which plays an important role in chlorophyll synthesis by affecting the transcription level of this enzyme to change chlorophyll content (Kong et al., 2016). In Ilex × attenuata 'Sunny Foster', the contents of ALA, protoIX, Mg-protoIX and pchlide in green-turned leaves were significantly increased, and the chlorophyll content was also significantly higher than that in normal leaves (Huang et al., 2021). Similarly, in *Camellia sinensis* cv. Baiye1, the contents of ALA, protoIX, Mg-protoIX and pchlide were higher in green leaves, and the chlorophyll content was also significantly higher than that in albino leaves (Wang et al., 2008). Similar results were obtained in this study, such as the contents of ALA, protoIX, Mg-ProtoIX and pchlide in *w-yl* were significantly lower than those in normal leaves ZK, indicating that the low chlorophyll content in *w-yl* may be due to the low content of chlorophyll precursors.

There are many kinds of leaf color mutations, and the genetic rules of different mutations vary greatly, which may be nuclear inheritance or cytoplasmic inheritance. For example, rice (Sun et al., 2017), maize (Wang et al., 2018) wheat (Jiang, 2018), cucumber (Gao et al., 2016), rape (Wang, 2014), tomato

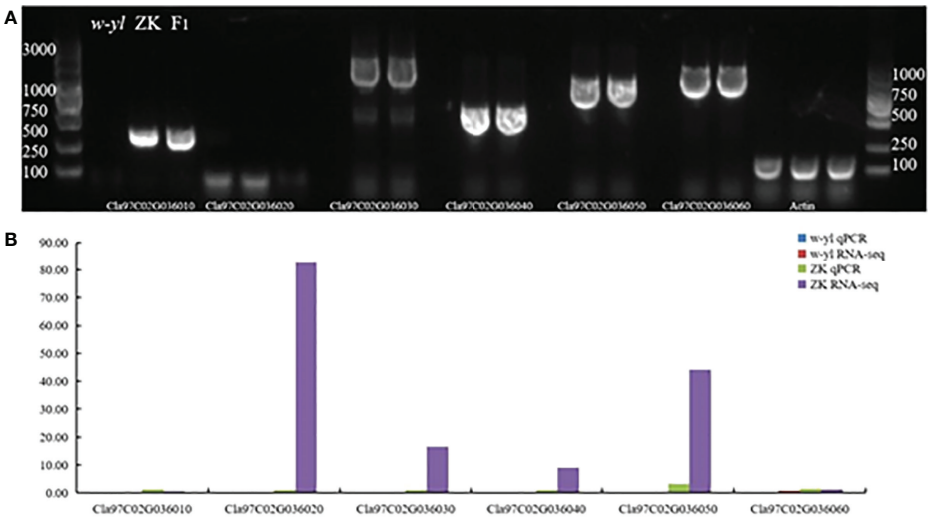


FIGURE 7
Amplification of candidate genes *Cla97C02G036010, Cla97C02G036020, Cla97C02G036030, Cla97C02G036040, Cla97C02G036050* and *Cla97C02G036060*. **(A)** Agarose gel electrophoresis analysis of candidate genes. **(B)** RNA-seq and qPCR analysis of candidate genes.

(Guo, 2017) and cabbage (Zhang, 2017) are controlled by single or two pairs of recessive nuclear genes. In watermelon, Zhang et al. (1996) proved that albino leaf color mutation was controlled by a pair of recessive alleles (*jaja*). Provvidenti (1994) found that the watermelon leaf color mottle mutation was controlled by a pair of recessive genes (*slv*), and the F_2 population showed a normal:mottled separation ratio of 3:1. Rhodes (1986) found that the post-green mutation was controlled by a recessive gene (*dgdg*). The data in this study indicated that *w-yl* is controlled by a pair of recessive nuclear genes. However, the results of mapping indicated that *w-yl* may have DNA fragment deletion compared to ZK, resulting in *Cla97C02G036010*, *Cla97C02G036030*, *Cla97C02G036040*, *Cla97C02G036050* and part of *Cla97C02G036060* in the interval between InD14,179,011 and InD16,396,362 loss the gene function. Chloroplast genome gene loss is relatively common in nature (Dong, 2012). Studies have shown that the most frequent microstructural changes in the chloroplast genome are insertions and deletions, and have a bias for deletions (Gao et al., 2010). In angiosperms, *rpl22, rpl23*, *rpl32*,

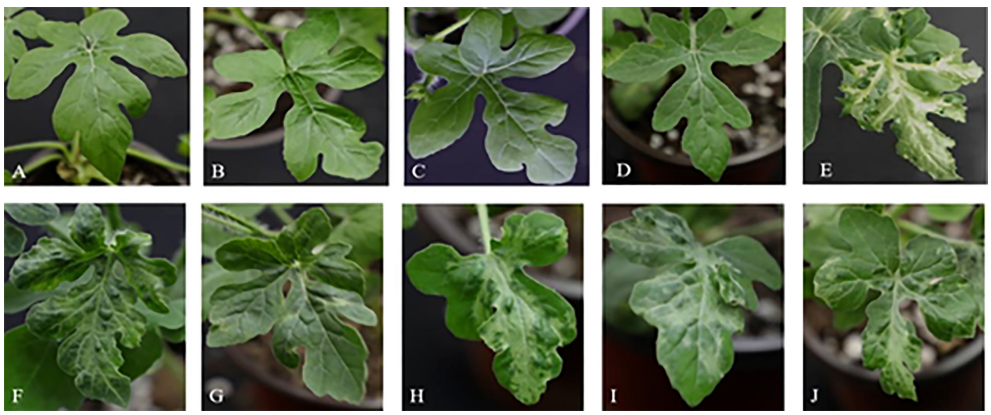


FIGURE 8
Leaf phenotypic analysis of candidate gene silencing. Phenotype of **(A)** blank control, **(B)** water control, **(C)** medium control, **(D)** blank vector control, **(E)** positive control, **(F)** silencing *Cla97C02G036010*, **(G)** silencing *Cla97C02G036030*, **(H)** silencing *Cla97C02G036040*, **(I)** silencing *Cla97C02G036050*, **(J)** silencing *Cla97C02G036060*.

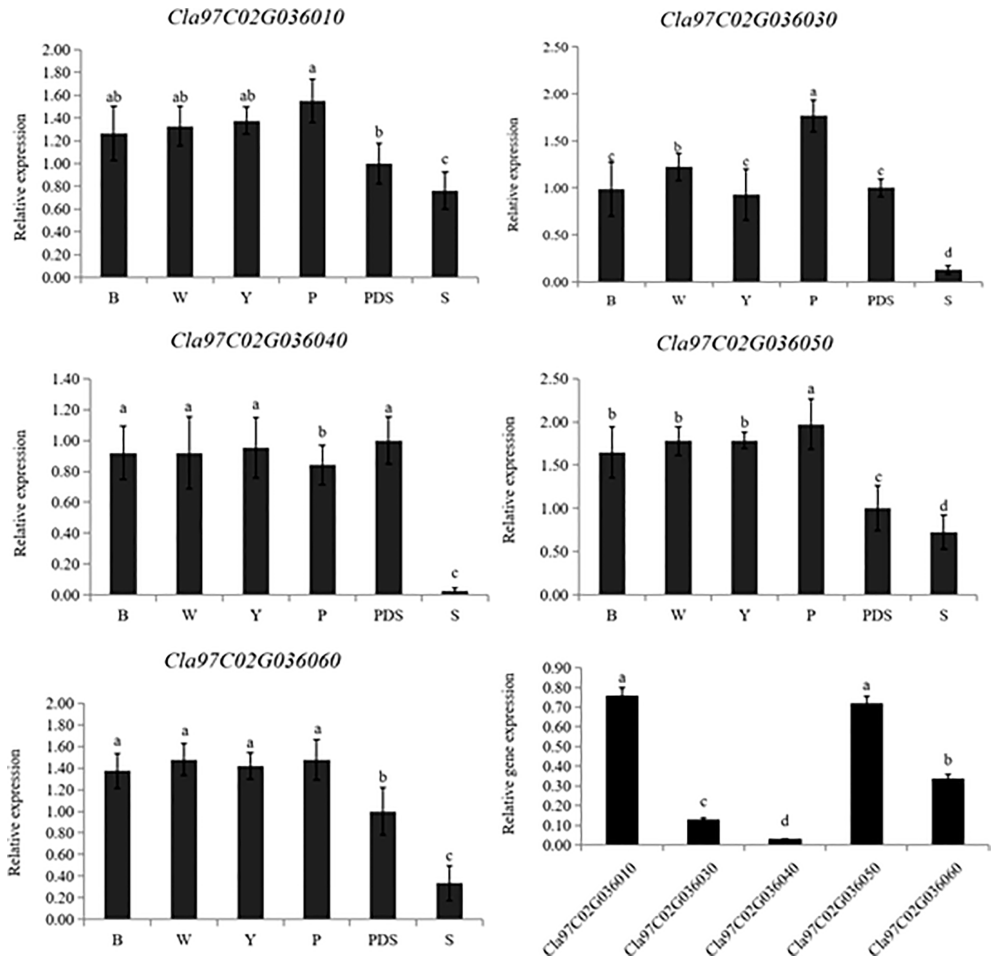


FIGURE 9
Expression levels of silenced genes. B, W, Y, P, PDS and S represents blank control, water control, YEP medium control, blank vector control, positive control and silenced genes, respectively. Small letters represent significant difference at P<0.05.

rpl33, *rps16*, *accD*, *psaI*, *ycf4*, *ycf1*, *ycf2 and infA* were lost in some taxa. Among them, *ycf1*, *ycf2* and *accD* genes were lost in the whole *Gramineae* (Guisinger et al., 2010) and some species in *Solanaceae* (Bruni et al., 2010).

In this study, the gene *Cla97C02G036060* encoded the protein Ycf2. NAD-malate dehydrogenase contained in the Ycf2/FtsHi complex is a key enzyme for ATP production in chloroplasts or non-photosynthetic plastids in the dark (Kikuchi et al., 2018), and is necessary for photosynthetic growth (Parker et al., 2016). The *Ycf2* gene is the largest plastid gene in angiosperms (Huang et al., 2010). It plays an important unknown function in higher plants and is indispensable (Drescher et al., 2000), which can response to biotic and abiotic stresses in plants (Durante et al., 2009) and improve plant cold tolerance (Bernardi et al., 2015). Gene *Cla97C02G036050* encoded a DnaJ-like B subfamily protein, which is a type of heat shock protein (Sun et al., 2018). Its homologous proteins can increase the activity of phytoene synthase in plastids (Zhou et al., 2015), participate in the process of white body differentiation into chloroplasts under light (Shimada et al., 2007), and protect plant photosystem II under heat stress (Wang and Luthe, 2003). For the gene *Cla97C02G036040*, it encoded a protein that containing the DUF679 domain. DUF (domain of unknown function) refers to a protein family with unknown functional domains, which is involved in regulating plant growth and development, plant defense mechanism and plant stress response in plants (Finn et al., 2016; Wang et al., 2022). In *Arabidopsis*, all members of DUF579 family can affect the development of xylan in plant cell wall hemicellulose (Temple et al., 2019), DUF761 is involved in regulating the growth and development of plant vegetative organs (Zhang et al., 2019), DUF642 protein is a specific

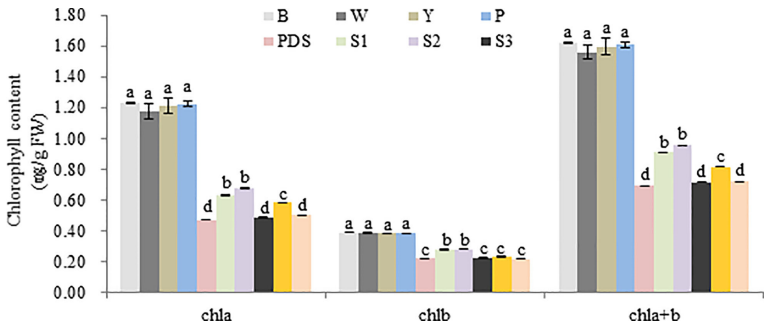


FIGURE 10
Chlorophyll content of different treatment group. B, W, Y, P, PDS, S1, S2, S3, S4,and S5 represents blank control, water control, YEP medium control, blank vector control, positive control, *Cla97C02G036010*, *Cla97C02G036030*, *Cla97C02G036040*, *Cla97C02G036050* and *Cla97C02G036060*, respectively. Small letters represent significant difference at P<0.05.

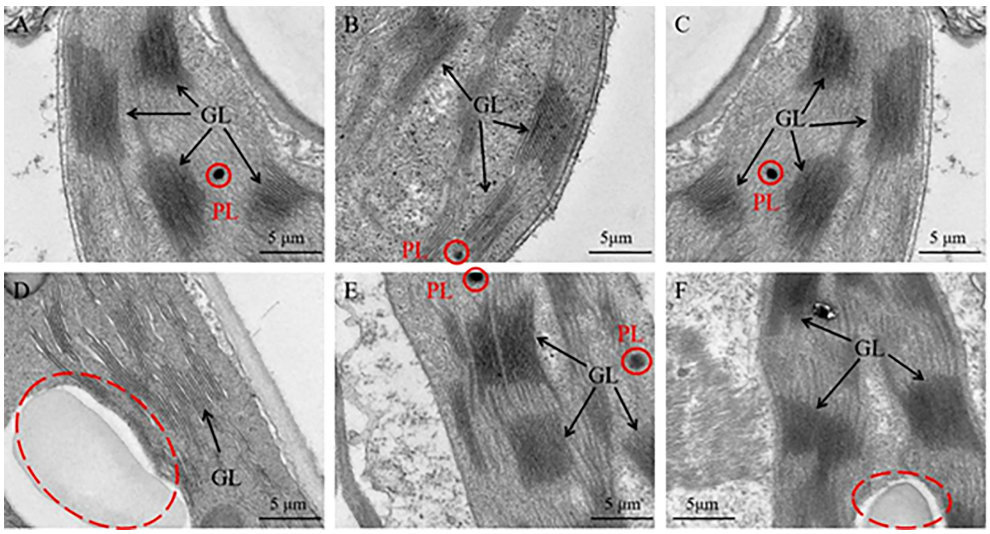


FIGURE 11
Ultrastructure of the chloroplast. **(A)** Chloroplast ultrastructure of blank control, Chloroplast ultrastructure of silencing gene **(B)***Cla97C02G036010*, **(C)** *Cla97C02G036030*, **(D)** *Cla97C02G036040*, **(E)** *Cla97C02G036050* and **(F)** *Cla97C02G036060*, respectively. GL represents grana lamella, PL represents plastoglobuli (red circle). The scale is 5 μm.

protein of seed plants, which is associated with cell wall synthesis (Vázquez-Lobo et al., 2012). In cucurbit crops, there are few studies on DUF domain, mainly focusing on the disease resistance of cucumber (Liu et al., 2010; Qin et al., 2018; Wang et al., 2018).

Data availability statement

The datasets presented in this study can be found inonline repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA872830.

Author contributions

JL, DS, YZ, and GY designed the experiments. YZ, GY, YW, GA, and WL provided experimental methods. YZ and GY performed the research and analyzed the data and wrote the manuscript. JL and DS reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (32102395), Natural Science Foundation of Xinjiang Uygur Autonomous Region (2020D01A136), Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2022-ZFRI), the China Agriculture Research System of MOF and MARA (CARS-25), and Joint Key Project of Agricultural Fine Variety in Henan Province (20220100001).

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.2022.1049114/full#supplementary-material

References

Awan, M. A., Konzak, C. F., Rutger, J. N., and Nilan, R. A. (1980). Mutagenic effects of sodium azide in rice. *Crop Sci.* 20, 663–668. doi: 10.2135/cropsci1980.0011183X002000050030x

Beale, S. L. (2005). Green genes gleaned. *Trends Plant Sci.* 10, 309–312. doi: 10.1016/j.tplants.2005.05.005

Bernardi, R., Bartolini, G., Petruccelli, R., Salvini, M., and Durantee, M. (2015). Modulated gene expression during the cold acclimation process in tolerant and sensitive clones of cultivar leccino (*Olea europaea* l.). *Plant Omics* 8, 405–411.

Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M., et al. (2010). Identification of poisonous plants by DNA barcoding approach. *Int. J. Legal Med.* 124, 595–603. doi: 10.1007/s00414-010-0447-3

Chen, T. T., Fu, W. M., Yu, J., Feng, B. H., Li, G. Y., Fu, G. F., et al. (2022). The photosynthetic characteristics of colored rice leaves and their relationship with antioxidant enzyme activities and anthocyanin content. in Chinese agricultural science. *Scientia Agricultura Sin.* 55, 467–478. doi: 10.3864/j.issn.0578-1752.2022.03.004

Chen, L. Y., He, L., Lai, J., He, S., Wu, X., and Zheng, Y. (2017). The variation of chlorophyll biosynthesis and the structure in different color leaves of bambusa multiplex 'Silverstripe'. *J. For. Environ.* 37, 385–391. doi: 10.13324/j.cnki.jfcf.2017.04.001

Chen, K., Li, Z., Cheng, M., Zhao, Y., Zhou, M., and Yang, H. (2018). Chloroplast ultrastructure and chlorophyll fluorescence characteristics of three cultivars of *Pseudosasa japonica*. *Chin. Bull. Bot.* 53, 509–518.DOI: 10.11983/CBB17115

Cui, H. R., Xia, Y. W., and Gao, M. W. (2001). Effects of temperature on leaf color and chlorophyll biosynthesis of rice mutant W1. *J. Nucl. Agric. Sci.* 15, 269–273. doi: 10.3969/j.issn.1000-8551.2001.05.003

Dei, M. (2010). Benzyladenine-induced stimulation of two components of chlorophyII formation in etiolated cucumber cotyledons. *Physiologia Plantarum* 62, 521–526. doi: 10.1111/j.1399-3054.1984.tb02793.x

Dong, W. (2012). *Chloroplast genome evolution of calycanthaceae andhighly variable chloroplast markers development of flowering plants* (Northeast Forestry University: Master).

Dong, H., Fei, G. L., Wu, C. Y., Wu, F. Q., Sun, Y. Y., Chen, M. J., et al. (2013). A rice virescent-yellow leaf mutant reveals new insights into the role and assembly of plastid caseinolytic protease in higher plants. *Plant Physiol.* 162, 1867–1880. doi: 10.1104/pp.113.217604

Drescher, A., Ruf, S., Calsa, T., Carrer, H., and Bock, R. (2000). The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22, 97–104. doi: 10.1046/j.1365-313x.2000.00722.x

Duan, Y. R., Gao, M. L., Guo, Y., Liang, X. X., Liu, X. J., Xu, H. G., et al. (2022). Map-based cloning and molecular marker development of watermelon fruit shape gene. *Scientia Agricultura Sin.* 55, 2812–2282. doi: 10.3864/j.issn.0578-1752.2022.14.011

Durante, M., Pieretti, M., and Bernardi, R. (2009). Chloropastic ycf2 gene expression in stressed plants. *53° Ital. Soc. Agric. Genet. Annu. Congres* 2, 45.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 44, 279–285. doi: 10.1093/nar/gkv1344

Franck, F., Sperling, U., Frick, G., Pochert, B., Van, C. B., Apel, K., et al. (2000). Regulation of etioplast pigment-protein complexes, inner membrane architecture, and protochlorophyllide alpha chemical heterogeneity by light-dependent NADPH:protochlorophyllide oxidoreductases a and b. *Plant Physiol.* 4, 1678–1696. doi: 10.1104/pp.124.4.1678

Gao, M., Hu, L., Li, Y., and Weng, Y. (2016). The chlorophyll-deficient golden leaf mutation in cucumber is due to a single nucleotide substitution in CsChlI for magnesium chelatase I subunit. *Theor. Appl. Genet.* 129, 1961–1973. doi: 10.1007/s00122-016-2752-9

Gao, P., Liu, S., Cui, H. N., Zhang, T. F., Wang, X. Z., Liu, H. Y., et al. (2020). Research progress of melon genomics,functional gene mapping and genetic engineering. *Acta Hortic. Sin.* 47, 1827–1844doi: 10.16420/j.issn.0513-353x.2020-0489.

Gao, L., Su, Y. J., and Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects. *J. Systematics Evol.* 48, 77–93. doi: 10.1111/j.1759-6831.2010.00071.x

Ge, Y. S., Yang, J., Wang, R. X., Hong, L., Jiang, W. H., Wu, Z. P., et al. (2022). Preliminary research on the green-revertible yellow leaf mutant S28-y in *Brassica napus*. *J. Yangzhou University(Agricultural Life Sci. Edition)* 43, 20–28. doi: 10.16872/j.cnki.1671-4652.2022.02.003

Gothandam, K., Kim, E., Cho, H., and Chung, Y. (2005). OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. *Plant Mol. Biol.* 58, 421–433. doi: 10.1007/s11103-005-5702-5

Guisinger, M., Chumley, T., Kuehl, J., Boore, J., and Jansen, R. (2010). Implications of the plastid genome sequence of typha (typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* 70, 149–160. doi: 10.1007/s00239-009-9317-3

Guo, L. J. (2017). *Genetic mapping of a variegated leaf (vg) mutant in tomato* (Huazhong Agricultural University: Master).

Guo, S. G., Zhao, S. J., Sun, H. H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4

Haileslassie, G. K. (2020). *Genetic and molecular mechanisms of delayed green leaf color and short internode length in watermelon (Citrullus lanatus)* (Chinese Academy of Agricultural Sciences).

Hibara, K., Obara, M., Hayashida, E., Ab, E. M., Ishimaru, T., Satoh, H., et al. (2009). The adaxialized leaf1 gene functions in leaf and embryonic pattern formation in rice. *Dev. Biol.* 334, 345–354. doi: 10.1016/j.ydbio.2009.07.042

Hou, Y., Zhu, Z. C., Zhu, N. N., Chen, K. N., Luan, F. S., and Wang, X. Z. (2016). Construction of EMS mutagenesis watermelon mutant library and phenotypic analysis. *Acta Botanica Boreali-Occidentalia Sin.* 36, 2411–2420. doi: 10.7606/j.issn.1000-4025.2016.12.2411

Huang, J. L., Sun, G. L., and Zhang, D. M. (2010). Molecular evolution and phylogeny of the angiosperm *ycf2* gene. *J. Systematics Evol.* 48, 240–248. doi: 10.1111/j.1759-6831.2010.00080.x

Huang, Q., Su, J., Zhou, P., Zhang, Q., and Zhang, M. (2021). Regulation of chlorophyll metabolism during leaf greening period of llex×attenuata 'Sunny foster'. *J. Northeast Forestry Univ.* 49, 51–54. doi: 10.13759/j.cnki.dlxb.2021.11.010

Huang, R., Wang, Y., Wang, P. R., Li, C. M., Xiao, F. L., Chen, N. G., et al. (2017). A single nucleotide mutation of IspFgene involved in the MEP pathway for isoprenoid biosynthesis causes yellow-green leaf phenotype in rice. *Plant Mol. Biol.* 96, 1–12.doi: 10.1007/s11103-017-0668-7

Hu, B. H., Wang, P., Du, A. P., Li, H., Wang, M. X., Bai, Y. L., et al. (2021). Characterization and gene mapping of *pyl3* mutant in rice. *He Nongxuebao (Journal Nucl. Agric. Sciences)* 35, 2696–2703. doi: 10.11869/j.issn.100-8551.2021.12.2696

Jiang, H. B. (2018). *Identification and gene mapping of chlorophyll- deficient mutant b23 in wheat* (Northwest Agriculture & Forestry University: Master).

Jiang, Q., Mei, J., Gong, X., Xu, J. L., Zhang, J. H., Teng, S., et al. (2014). Importance of the rice TCD9 encoding α subunit of chaperonin protein 60 (Cpn60α) for the chloroplast development during the early leaf stage. *Plant Sci.* 56, 400–410. doi: 10.1016/j.plantsci.2013.11.003

Jin, F. L., Chen, L. Q., Zhang, J. L., Wang, J. J., Li, H. Y., Han, Y. H., et al. (2021). Effects of different ilumination intensities on yellow leaf mutant seedlings of broomcorn millet. *J. Shanxi Agric. University(Natural Sci. Edition)* 41, 26–34. doi: 10.13842/j.cnki.issn1671-8151.202104035

Kenneth, J. L., and Thomas, D. S. (2002). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., et al. (2018). A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* 30, 2677–2703. doi: 10.1105/tpc.18.00357

Kong, W. Y., Yu, X. W., Chen, H. Y., Liu, L. L., Xiao, Y. J., Wang, Y. L., et al. (2016). The catalytic subunit of magnesium-protoporphyrin IX monomethyl ester cyclase forms a chloroplast complex to regulate chlorophyll biosynthesis in rice. *Plant Mol. Biol.* 92, 177–191. doi: 10.1007/s11103-016-0513-4

Lai, Y., Fu, Q. S., Lv, J. C., Zhou, M., He, M., Xu, W. P., et al. (2018). Analysis of physiological characteristics and chloroplast ultrastructure of a new leaf color mutant in melon. *J. Sichuan Agric. Univ.* 36, 372–379. doi: 10.16036/j.issn.1000-2650.2018.03.015

Lee, H. J., Ball, M. D., Parham, R., and Rebeiz, C. A. (1992). Chloroplast biogenesis 65:enzymic conversion of protoporphyrin ix to mg-protoporphyrin ix in a subplastidic membrane fraction of cucumber etiochloroplasts. *Plant Physiol.* 99, 1134–1140. doi: 10.1104/pp.99.3.1134

Li, Y. (2016). *Physiological characteristics and genetic analysis of leaf color mutant in cucumber* (Sichuan Agricultural University).

Li, R. Q., Jiang, M., Huang, J. Z., Møller, I. M., and Shu, Q. Y. (2021b). Mutations of the genomes *uncoupled4* gene cause ros accumulation and repress expression of peroxidase genes in rice. *Front. Plant Sci.* 12, 682453–682453. doi: 10.3389/fpls.2021.682453

Liu, M. (2019). *Construction of a VIGS vector based on cucumber green mottle mosaic virus* (Chinese Academy of Agricultural Sciences: Master).

Liu, J., Tian, H. L., Wang, Y. H., and Guo, A. G. (2010). A novel antibacterial protein gene GNK2-1 in cucumber and its resistance to fusarium wilt. *Botanical J.* 45, 411–418. doi: 10.3969/j.issn.1674-3466.2010.04.003

Li, C., Wang, J. W., Hu, Z. Y., Xia, Y. Y., Huang, Q., Yu, T., et al. (2021a). A valine residue deletion in ZmSig2A, a sigma factor, accounts for a revertible leaf-color mutation in maize. *Crop J.* 9, 1330–1343. doi: 10.1016/j.cj.2021.01.005

Li, S. Z., Yang, W. Z., and Chen, R. M. (2018). An overview on yellow green leaf mutants in rice. *Biotechnol. Bull.* 34, 15–21. doi: 10.13560/j.cnki.biotech.bull.1985.2018-0465

Ma, S. W., and Zhang, L. (1999). Discovery of mutant strains carrying watermelon albino lethal gene. *Chin. watermelon melon* 4, 22. doi: 10.16861/j.cnki.zggc.1999.04.009

Parker, N., Wang, Y., and Meinke, D. (2016). Analysis of *Arabidopsis* accessions hypersensitive to a loss of chloroplast translation. *Plant Physiol.* 172, 1862–1875. doi: 10.1104/pp.16.01291

Provvidenti, R. (1994). Inheritance of a partial chlorophyll deficiency in watermelon activated by low temperatures at the seedling stage. *Horticulture Sci.* 29, 1062–1063. doi: 10.21273/HORTSCI.29.9.1062

Qin, Z. E., Wang, Y. T., Liu, D., Xin, M., and Zhou, X. Y. (2018). CsCBS clone of cucumber and preliminary verification of its resistance to downy mildew and corynebacterium leaf spot. *J. Northeast Agric. Univ.* 49, 39–47. doi: 10.19720/j.cnki.issn.1005-9369.2018.02.005

Rebeizjames, C. A., Mattheis, R., Smith, B. B., Rebeiz, C. C., and Dayton, D. F. (1975). Chloroplast biogenesis: Biosynthesis and accumulation of mg-protoporphyrin IX monoester and longer wavelength metalloporphyrins by greening cotyledons. *Arch. Biochem. Biophysics* 167, 446–465 10.1016/0003-9861(75)90408-7

Ren, Y. C., Zhu, Y. C., Sun, D. X., Deng, Y., An, G. L., Li, W. H., et al. (2019). Physiological characteristic analysis of a leaf-yellowing mutant in watermelon. *J. Fruit Sci.* 37, 565–573. doi: 10.13925/j.cnki.gsxb.20190508

Rhodes, B. (1986). Genes affecting foliage color in watermelon. *J. Heredity* 77, 134–135. doi: 10.1093/oxfordjournals.jhered.a110190

Sakuraba, Y., Rahman, M. L., Cho, S. H., Kim, Y. S., Koh, H. J., Yoo, S. C., et al. (2013). The rice faded green leaf locus encodes protochlorophyllide oxidoreductase b and is essential for chlorophyll synthesis under high light conditions. *Plant J.* 74, 122–133. doi: 10.1111/tpj.12110

Shao, Q. (2013). *Characterization and proteomics of a novel xantha mutant in muskmelon* (Northeast Agricultural University).

Shao, Q., Yu, Z. Y., Li, X. G., Li, W., and Gao, Y. (2013). Studies on internal physiological and biochemical changes of xantha mutant in melon leaves. *China Vegetables* 14, 59–65. doi: 10.3969/j.issn.1000-6346.2013.14.012

Shimada, H., Mochizuki, M., Ogura, K., Froehlich, J. E., Osteryoung, K. W., Shirano, Y., et al. (2007). *Arabidopsis* cotyledon-specific chloroplast biogenesis factor CYO1 is a protein disulfide isomerase. *Plant Cell* 19, 3157–3169. doi: 10.1105/tpc.107.051714

Sugliani, M., Abdelkefi, H., Ke, H., Bouveret, E., Robaglia, C., Caffarri, S., et al. (2016). An ancient bacterial signaling pathway regulates chloroplast function to influence growth and development in *Arabidopsis*. *Plant Cell* 28, 661–679. doi: 10.1105/tpc.16.00045

Sun, L., Lin, T., Wang, Y., Niu, M., Hu, T., Liu, S., et al. (2017). Phenotypic analysis and gene mapping of a white stripe mutant st13 in rice. *Chin. J. Rice Sci.* 31, 355–363. doi: 10.16819/j.1001-7216.2017.6169

Sun, J. L., Tian, Y. Y., Lian, Q. C., and Liu, J. X. (2020). Mutation of DELAYED GREENING impairs chloroplast RNA editing at elevated ambient temperature in *Arabidopsis*. *J. Genet. Genomics* 47, 200+202–212.

Sun, C. C., Yan, F., and Chen, J. P. (2018). The role of *N. benthamiana* DNA J-like protein in the infection of turnip mosaic virus. *Zhejiang Agric. J.* 30, 2056–2064. doi: 10.3969/j.issn.1004-1524.2018.12.10

Su, J., Shi, W., Yang, Y., Wang, X., Ding, Y., and Lin, S. (2020). Comparison of leaf color and pigment content and observation of leaf structure at different growth stages from six bamboo species. *Scientia Silvae Sinicae* 56, 194–203. doi: 10.11707/j.1001-7488.20200720

Tan, J., Zhang, T., Xia, S. S., Yan, M., Li, F. F., Sang, X. C., et al. (2019). Fine mapping of anovel yellowgreen leaf 14 (ygl14) mutant in rice. *Euphytica* 215, 1–11. doi: 10.1007/s10681-019-2424-3

Temple, H., Mortimer, J. C., Tryfona, T., Yu, X., Lopez-Hernandez, F., Sorieul, M., et al. (2019). Two members of the DUF579 family are responsible for arabinogalactan methylation in *Arabidopsis*. *Plant Direct* 3, e00117. doi: 10.1002/pld3.117

Vázquez-Lobo, A., Roujol, D., Zuiga-Sánchez, E., Albenne, C., Piñero, D., Debuen, A. G., et al. (2012). The highly conserved spermatophyte cell wall DUF642 protein family: Phylogeny and first evidence of interaction with cell wall polysaccharides *in vitro*. *Mol. Phylogenet. Evol.* 63, 510–520. doi: 10.1016/j.ympev.2012.02.001

Wang, J. (2014). *Identification seed purity of "shaanyou 803" in brassca napus l. and molecular markersof one gene underlying chlorophiyll-deficit trait in b. juncea* (Northwest Agriculture & Forestry University: Master).

Wang, F., Duan, S., Li, T., Wang, N., and Tao, Y. (2018). Fine mapping and candidate gene analysis of leaf color mutant in maize. *J. Plant Genet. Resour.* 19, 1205–1209. doi: 10.13430/j.cnki.jpgr.20180326001

Wang, X. R., Hu, Q., Du, X. Z., and Feng., S. (2022). Identification of rice DUF642 family genes and their expression analysis under abiotic stress. *J. Hubei Univ. (Natural Sci. Edition)* 44, 14–23. doi: 10.3969/j.issn.1000-2375.2021.00.010

Wang, D., and Luthe, D. S. (2003). Heat sensitivity in a bentgrass variant. failure to accumulate a chloroplast heat shock protein isoform implicated in heat tolerance. *Plant Physiol.* 133, 319–327. doi: 10.1104/pp.102.018309

Wang, J. M., Ma, S. W., Shang, J. L., and Cheng, S. H. (2011). Leaf colour mutants in cucurbits and their research progress. *J. Anhui Agric. Sci.* 39, 16039–16040+16116. doi: 10.3969/j.issn.0517-6611.2011.26.082

Wang, F. C., and Wang, H. B. (1997). A briefing on the new mutant of watermelon "Yellow bud". *Chin. watermelon melon* 3, 15–16.

Wang, J. Y., Wang, Z. P., and Duan, X. K. (2019). Discovery and genetic analysis of yellow markers in melon buds. *China Vegetables* 32, 15–17.

Wang, P. R., Zhang, F. T., Gao, J. X., Sun, X., and Deng, X. (2009). An overview of chlorophyll biosynthesis in higher plants. *Acta Botanica Boreali-Occidentalia Sin.* 29, 629–636. doi: 10.3321/j.issn:1000-4025.2009.03.032

Wang, X. C., Zhao, L., Yao, M., Chen, L., and Yang, Y. (2008). Preliminary study on gene expression differences between normal leaves and albino leaves of anji baicha (*Camellia sinensis* cv.Baiye1). *J. Tea Sci.* 1, 50–55.

Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L., et al. (2019). Genome of 'Charleston gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. national plant germplasm system watermelon collection. *Plant Biotechnol. J.* 153, 994–1003. doi: 10.1111/pbi.13136

Xiao, H. G., Yang, H. W., Rao, Y., Yang, B., and Zhu, Y. (2013). Photosynthetic characteristics and chlorophyll fluorescence kinetic parameters analyses of chlorophyll-reduced mutant in *Brassica napus* l. *Acta Agronomica Sin.* 39, 520–529. doi: 10.3724/SP.J.1006.2013.00520

Xiong, L. R., Du, H., Zhang, K. Y., Lv, D., He, H. L., Pan, J. S., et al. (2020). A mutation in CsYL2.1 encoding a plastid isoform of triose phosphate isomerase leads to yellow leaf 2.1 (*yl2.1*) in cucumber (*Cucumis sativus* l.). *Int. J. Mol. Sci.* 22, 322. doi: 10.3390/ijms22010322

Xu, M., Gao, M. L., Guo, Y., Bao, X. P., Liu, X. J., Liu, J. X., et al. (2022). Photosynthetic characteristics of virescent mutant in watermelon. *J. Northwest A F University(Natural Sci. Edition)* 50, 91–96+106. doi: 10.13207/j.cnki.jnwafu.2022.03.012

Xu, J., Guo, Z. X., jiang, X. X., Ahammedd, G. J., and Zhou, Y. H. (2021). Light regulation of horticultural crop nutrient uptake and utilization. *Horticult Plant J.* 5, 367–379. doi: 10.1016/j.hpj.2021.01.005

Yang, Q. S., He, H., Li, H. Y., Tian, H., Zhang, J. J., Zhai, L. G., et al. (2011). NOA1 functions in a temperature-dependent manner to regulate chlorophyll biosynthesis and rubisco formation in rice. *PloS One* 6, e20015. doi: 10.1371/journal.pone.0020015

Yang, J., Shi, S. L., Ji, X. H., Zhao, L. Q., and Xu, C. Q. (2018). Efect of low temperature stres on physiological indexes of eight species color-leafed trees. *Beifang Yuanyi (Northern Horticulture)* 5, 106–110. doi: 10.11937/bfyy.20172469

Yang, C., Zhang, Y. Y., Fang, Z. Y., Liu, Y. M., Yang, L. M., Zhuang, M., et al. (2014). Photosynthetic physiological characteristics and chloroplast ultrastructure of yellow leaf mutant YL-1 in cabbage. *Acta Hortic. Sin.* 41, 1133–1144. doi: 10.16420/j.issn.0513-353x.2014.06.028

Yao, J. G., Zhang, H., Xu, X. Y., Zhang, L. L., and Li, J. F. (2010). Studies on weak light tolerance of tomato leaf color mutant. *China Vegetables* 4, 31–35. doi: CNKI:SUN:ZGSC.0.2010-04-010

Yuan, G. P., Chun, C. P., Peng, L. Z., Huang, Z. J., Huang, T., Yang, H., et al. (2017). A study on the difference of 'Newhall' navel orange and its sport 'Longhuihong' navel orange. *J. Fruit Sci.* 34, 1117–1124. doi: 10.13925/j.cnki.gsxb.20170096

Yuan, Y. H., Fan, J. L., Yang, W. Z., and Chen, R. M. (2022b). Study on photosynthetic characteristics of 3 maize yellow leaf mutants. *Adv. Biotechnol.* 12, 75–82. doi: 10.19586/j.2095-2341.2021.0097

Yuan, G., Sun, D., An, G., Li, W., Si, W., Liu, J., et al. (2022a). Transcriptomic and metabolomic analysis of the effects of exogenous trehalose on salt tolerance in watermelon (*Citrullus lanatus*). *Cells* 11, 2338. doi: 10.3390/cells11152338

Zhang, K. (2017). *Inheritance and gene mapping of a yellow leaf mutant pylm in pakchoi* (Shenyang Agricultural University).

Zhang, H., Li, J. J., Yoo, J. H., Yoo, S. C., Cho, S. H., Koh, H. J., et al. (2006). Rice chlorina-1 and chlorina-9 encode ChlD and ChlI subunits of mg-chelatase, a key enzyme for chlorophyll synthesis and chloroplast development. *Plant Mol. Biol.* 62, 325–337. doi: 10.1007/s11103-006-9024-z

Zhang, X. P., Rhodes, B. B., Baird, W. V., Skorupska, H. T., and Bridges, W. (1996). Development of genic male-sterile watermelon lines with delayed-green seedling marker. *Hortic. Sci.* 31, 123–126. doi: 10.21273/HORTSCI.31.1.123

Zhang, Z. P., Wang, J. Y., Xing, G. M., Li, M. L., and Li, S. (2022). Integrating physiology, genetics, and transcriptome to decipher a new thermo-sensitive and light-sensitive virescent leaf gene mutant in cucumber. *Front. Plant Sci.* 13, 972620–972620. doi: 10.3389/fpls.2022.972620

Zhang, Y., Zhang, F., and Huang, X. Z. (2019). Characterization of an arabidopsis thaliana DUF761-containing protein with a potential role in development and defense responses. *Theor. Exp. Plant Physiol.* 31, 303–316. doi: 10.1007/s40626-019-00146-w

Zhang, T. Y., Zhou, C. L., Liu, X., Sun, A. L., Cao, P. H., Nguye, T., et al. (2017). Phenotypes and gene mapping of a thermo-sensitive yellow leaf mutant of rice. *Acta Agronomica Sin.* 43, 1426–1433. doi: 10.3724/SP.J.1006.2017.01426

Zhou, X., Welsch, R., Yang, Y., Lvarez, D., Riediger, M., Yuan, H., et al. (2015). *Arabidopsis* OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci. United States America* 112, 3558–3563. DOI: 10.1073/pnas.1420831112

Zhu, H. Y., Zhang, K. G., Song, P. Y., Zhang, Y., Wang, X. J., Hu, J. B., et al. (2019). Genetic analysis and gene mapping of yellow-green leaf in melon(*Cucumis melo* l.). *J. Henan Agric. Univ.* 53, 16–21. doi: 10.16445/j.cnki.1000-2340.20191025.001

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpls.2022.1047090

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Rong Zhou,
Aarhus University, Denmark
Xingwang Liu,
China Agricultural University, China
Junjun Shen,
Northwest A&F University, China

*CORRESPONDENCE
Yu'e Lin
linyue@gdaas.cn
Biao Jiang
jiangbiao@gdaas.cn

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

RECEIVED 17 September 2022
ACCEPTED 03 October 2022
PUBLISHED 20 October 2022

CITATION
Yan J, Liu B, Cao Z, Chen L, Liang Z, Wang M, Liu W, Lin Y and Jiang B (2022) Cytological, genetic and transcriptomic characterization of a cucumber albino mutant. *Front. Plant Sci.* 13:1047090. doi: 10.3389/fpls.2022.1047090

Cytological, genetic and transcriptomic characterization of a cucumber albino mutant

Jinqiang Yan[1,2], Bin Liu[3], Zhenqiang Cao[1], Lin Chen[1,2], Zhaojun Liang[1], Min Wang[1,2], Wenrui Liu[1,2], Yu'e Lin[1]* and Biao Jiang[1,2]*

[1]Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [2]Guangdong Key Laboratory for New Technology Research of Vegetables, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [3]Hami-melon Research Center, Xinjiang Academy of Agricultural Sciences, Urumqi, China

Photosynthesis, a fundamental process for plant growth and development, is dependent on chloroplast formation and chlorophyll synthesis. Severe disruption of chloroplast structure results in albinism of higher plants. In the present study, we report a cucumber albino *alc* mutant that presented white cotyledons under normal light conditions and was unable to produce first true leaf. Meanwhile, *alc* mutant could grow creamy green cotyledons under dim light conditions but died after exposure to normal light irradiation. No chlorophyll and carotenoid were detected in the *alc* mutant grown under normal light conditions. Using transmission electron microscopy, impaired chloroplasts were observed in this mutant. The genetic analysis indicated that the albino phenotype was recessively controlled by a single locus. Comparative transcriptomic analysis between the *alc* mutant and wild type revealed that genes involved in chlorophyll metabolism and the methylerythritol 4-phosphate pathway were affected in the *alc* mutant. In addition, three genes involved in chloroplast development, including two *FtsH* genes and one *PPR* gene, were found to have negligible expression in this mutant. The quality of RNA sequencing results was further confirmed by real-time quantitative PCR analysis. We also examined 12 homologous genes from *alc* mutant in other plant species, but no genetic variation in the coding sequences of these genes was found between *alc* mutant and wild type. Taken together, we characterized a cucumber albino mutant with albinism phenotype caused by chloroplast development deficiency and this mutant can pave way for future studies on plastid development.

KEYWORDS

albino mutant, cucumber, recessive, chloroplast deficiency, transcriptome

Introduction

Chloroplasts, which are DNA-containing organelles play crucial roles in attuning plant development and plant interaction with environmental cues. At the time of illumination, the chloroplast develops from proplastids *via* the process of photomorphogenesis (Pogson et al., 2015; Chan et al., 2016). The chloroplast is the site of photosynthesis and production of hormones (e.g., abscisic acid, jasmonic acid, and salicylic acid, and other major metabolites (Unlu et al., 2014). Its abnormal development or accumulation of pigments inside itself could affect photosynthesis and further disrupt plant growth and biomass yield (Wang et al., 2016; Shi et al., 2017; Xiong et al., 2017; Gotoh et al., 2018).

Leaf color mutation has been widely reported in many plant species. In most of the chlorophyll-less mutants and chlorophyll-deficient mutants, a sudden increase in the production of reactive oxygen species (ROS) was detected after exposure to light conditions (Sakowska et al., 2018; Li et al., 2019). The ROS accumulation induced by excessive light could effectuate oxidative damage in plants, resulting in leaf bleaching or leading to plant death. Natural or induced albino mutants were frequently identified and characterized among different kinds of leaf color mutations, especially in *Arabidopsis* and rice. For example, some formation of albino mutants was affected by environmental factors and conditionally green-revertible. Disruption of *OsABCI8* resulted in the development of albino leaves in rice under continuous rainy days; nevertheless, the leaves gradually turned green following rainy days (Zeng et al., 2017). Mutations in the gene *OsTCD5* encoding a monooxygenase, or *OsTCD11* encoding the ribosomal small subunit protein S6 in chloroplasts (RPS6) resulted in a temperature-sensitive albino mutant; the leaves displayed albinism at low temperatures but turned green at high temperatures (Wang et al., 2016). Mutation in gene *FLN2* encoding fructokinase-like protein2 shows opposite phenotype; the *fln2* mutant is albino at high temperatures (Qiu et al., 2018). There are other mutants that only exhibit albino phenotype at certain development stages. Disruption of a pentatricopeptide repeat protein causes an albino phenotype during the seedling stage but the plants are able to turn green during plant growth and development (Su et al., 2012). A mutation in SEEDLING PLASTID DEVELOPMENT1 resulted in albino cotyledons, but these plants appeared similar to the wild type plants once the initial true leaves developed and the seedlings were transferred to the soil (Ruppel et al., 2011).

Somatic albino mutants were also described with expression of genes involved in chlorophyll biosynthesis and chloroplast development affected in mutated leaves (Ma et al., 2018; Lu et al., 2019). Some lethal mutations are caused by deficiency in chloroplast development. Loss of function of *DXS1* leads to an albino phenotype in tomato with premature lethality performance (Garcia-Alcazar et al., 2017). A single-nucleotide mutation in the plastid ribosomal protein produces abnormal chloroplasts and causes seedling lethality in rice (Zhao et al., 2016).

Cucumber (*Cucumis sativus* L.), belonging to the Cucurbitacea family is an important vegetable crop worldwide. Cucumber albino mutation was only reported by Iida and Amano in 1991 (Iida and Amano, 1991), which was induced by irradiation. However, no further studies have been carried out since then. In the present study, we reported a spontaneous mutation of an albino mutant from the cucumber inbred line "g32", which exhibited white cotyledons and hypocotyl, and died before developing any true leaves. We aim to decipher the potential mechanism of albinism formation in this cucumber albino mutant *via* cytologic, genetic and transcriptomica characterization.

Materials and methods

Plant materials

Cucumber inbred line g32 was used in this study. The inbred line was provided by Vegetable Research Institute, Guangdong Academy of Agricultural Sciences Guangzhou, China. Seeds from a self-pollinated g32 cucumber fruit were soaked in water for 4 h and then kept in the incubator with moderate humidity at 28 °C for germination. Thereafter, germinated seeds were planted in a plug tray either under artificial light irradiation (LED light model, defined as normal light) or dim light in the greenhouse of Vegetable Research Institute, Guangdong Academy of Agricultural Sciences. Dim light treatment was performed in a black plastic bag covered homemade growth chamber. Light intensity of normal light and of dim light was 8000 LUX and 35 LUX, respectively. Seven-day-old seedlings of wild type and *alc* mutant were used for phenotypic evaluation, fluorescence microscopy, transmission electron microscopy analysis, and high-throughput RNA sequencing.

Pigment content measurement

To determine the chlorophyll a (Chla), chlorophyll b (Chlb), total chlorophyll (Chl), and carotenoid (Car) contents, wild type and *alc* mutant seedlings grown under normal and dim light conditions were compared. A total of 0.1 g cotyledon from each sample was cut into small pieces and transferred into 10 mL 80% (v/v) acetone and kept in dark until the tissue turned white. Each sample was analyzed in three biological replicates. For each sample, the absorbance was measured at 663, 645, and 470 nm thrice, respectively. Concentrations of Chla, Chlb, Chl, and Car were calculated as described previously (Lichtenthaler, 1987).

Fluorescence microscopy

For fluorescence microscopy, the abaxial epidermis of cotyledons was used. Chloroplast autofluorescence (red) was captured under Zeiss LSM710 (Germany) confocal microscope with the following settings: excitation at 633 nm, emission at 647–721 nm. Data were analyzed using software ZEN (2010).

Transmission electron microscopy(TEM)

Cotyledons of wild type and *alc* mutant seedlings grown under dim light conditions and normal light irradiation were analyzed by TEM. All the cotyledon samples were cut into 1–2 mm^2 sections and fixed in 2.5% glutaraldehyde and 4% paraformaldehyde in phosphate buffer (pH 6.8–7.2) under vacuum for 3 h. After washing with phosphate buffer the samples were fixed in 1% osmium tetroxide (OsO_4) for 3 h and again washed with phosphate buffer. The samples were dehydrated through a series of ethanol concentration. The samples were infiltrated with an increasing ratio of Spurr's resin dilutions [25%, 50%, 75%, and 100% (v/v)] to substitute ethanol, and finally embedded in Spurr's resin. After cutting, the sections were viewed under a HitachiH-7700 (Hitachi) transmission electron microscope.

RNA extraction

Total RNA was extracted from cotyledons of 7-day-old wild type and *alc* mutant seedlings using Trizol Kit (Promega, USA) according to the manufacturer's instructions. Extracted RNA was treated with RNase-free DNase I (TaKaRa, Japan) to remove residual DNA. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

cDNA library construction, high-throughput sequencing and mapping

Cotyledons of wild type and *alc* mutant seedlings grown under normal light conditions were collected for high-throughput sequencing. Each sample was analyzed in three biological replicates. A total of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. Library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated. Thereafter, reads with adaptors, reads with unknown bases, as well as low quality reads were removed to generate clean reads. The remaining high-quality clean reads were mapped to Cucumber (Chinese Long) Reference Genome v2 (http://www.cucurbitgenomics.org/organism/2).

Quantification of gene expression, GO, and KEGG pathway enrichment analysis

The mapped reads of each sample were assembled using StringTie (v1.3.3b) 17 in a reference-based approach, and featureCounts v1.5.0-p3 18 was used to count the reads numbers mapped to each gene. Fragments Per Kilobase of transcript sequence per Millions base pairs (FPKM) of each gene was calculated based on the gene length and read count mapped to this gene. Differential expression analysis of two groups was performed using the DESeq2 R package (1.16.1) 19. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value (padj) <0.05 and |log2 (FoldChange)| > 2 were assigned as differentially expressed genes (DEGs). To functionally annotate the DEGs, Gene Ontology (GO, http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) annotation of the unigenes were analyzed using clusterProfiler R package.

Quantitative real-time PCR validation

To confirm RNA-seq results, 12 DEGs were selected for qRT-PCR validation. First strand cDNA synthesis was performed using TransScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (Transgen, China) with 1 μg of RNA used for high-throughput sequencing. Quantitative RT-PCR was carried out using 0.2 μg cDNA using PerfectStart Green qPCR Supermix (Transgen, China) according to the manufacturer's instructions. Reactions were performed and analyzed on CFX96 Real-Time PCR Detection System. Three biological replicates and three technical replicates per sample were performed for each gene. Gene expression was normalized against α-TUBULIN *(TUA)* gene (Liu et al., 2016). Primers used are listed in Supplementary Table 1.

Results

Phenotypic characterization of a cucumber *alc* mutant

We observed a few naturally occurring albino seedlings during the reproduction of cucumber inbred line "g32", which is a southern China type cucumber. The cotyledons of these

seedlings were small and entirely white, with short and white hypocotyl and short primary root (Figure 1). The mutant was named as *alc* (albino cotyledon) thereafter. The mutants dried out and died in a few days after emerging from the substrate without growing any true leaves.

It was noteworthy that occasionally cotyledons of a few *alc* mutants that were still inside the shell showed subtle greenish; this led us to propose that the occurrence of greenish cotyledons in *alc* mutants may be affected by light. Therefore, we performed three sets of experiments to validate this hypothesis (Figure 2A). In the first experiment, all the seeds were grown under dim light conditions at all times. After emerging from the substrate, *alc* mutants presented creamy green cotyledons with complete white hypocotyl, while wild type seedlings showed yellowish-green cotyledons and hypocotyl (Figure 2B). In the continually dim light environment, both *alc* mutant and wild-type plants spindled and died without growing first true leaf. In the second experiment, after the seeds emerged from the substrate, they were first allowed to grow under normal light conditions until we were able to distinguish between *alc* mutant and wild-type plants. We then transferred them to dim light conditions and observed that the cotyledons of *alc* mutants remained whitish, without turning to cream green (Figures 2C1, C2). In the last experiment, the seeds were grown under dim light conditions until seedlings emerged from substrate, and then they were exposed to normal light condition (Figure 2A3). After 5 h of light exposure, the green color in the cotyledon of *alc* mutant started to degrade (Figure 2D1). The cotyledon shrunk and dried out after exposure to light for 30 h (Figure 2D2). Therefore, we conclude that *alc* is a light sensitive albino mutant. Normal light irradiation is a lethal factor for *alc* mutant and, the damage caused by light is irreversible.

Most of the albino phenotypes in other plants lack chlorophyll and carotenoids. Thus, we examined the chlorophyll and carotenoid content of cotyledons from both wild type and *alc* mutant seedlings that grew under normal light and dim light conditions. Fluorescence microscopy revealed that chlorophyll fluorescence in wildtype was more intense than that observed in *alc* mutants grown under dim light conditions (Figure 2E; presented in red). As expected, no chlorophyll fluorescence was detected in *alc* mutants that grew under normal light conditions. Chlorophyll content was measured based on the fluorescence intensity (Figure 2G1). No carotenoid was detected in *alc* mutants grown under normal light conditions; however, a small amount, 0.03 mg/g (fresh weight), was detected in *alc* mutants grown under dim light conditions (Figure 2G2).

Since most of the chlorophyll content in plants are in chloroplast, we further investigated the chloroplast ultrastructure in the cotyledons of wild type and *alc* mutant seedlings that grew under both light and dim light conditions using TEM. In the cotyledons of wild type seedlings, we observed numerous well-developed, crescent-shaped chloroplasts with stroma thylakoids, grana thylakoids, starch granules, and plastoglobuli within the membranes Figures 2F1, F4, F7). In contrast, the chloroplasts in the *alc* mutant decreased dramatically in number and showed abnormal shapes that lacked stroma and grana thylakoids, but contained osmiophilic plastoglobulis in the inner membrane system (Figures 2F2, F5, F8). The *alc* mutants that grew under dim light conditions comprised stroma thylakoid as well as osmiophilic plastoglobulis (Figures 2F3, F6, F9).

To summarize, the above results indicated that the chloroplast development was impaired in *alc* mutants grown under normal light conditions during seedling development. Moreover, normal light may be lethal to the *alc* mutant because it interrupts thylakoid biogenesis, as observed by the presence of thylakoid in the *alc* mutant grown in dim light conditions but not in those grown under normal light conditions.

Inheritance model of the *alc* mutant

As the *alc* mutant died in the seedling stage, we could not obtain homozygous seeds. Therefore, we considered the mutant parental line "g32" as F_1 and used its self-pollinated seeds to investigate the inheritance pattern of the albino phenotype. We planted a total of 123 seeds from the self-pollinated "g32" cucumber and put them under normal light conditions, of which 116 seeds germinated successfully (germination rate 94.3%). Thirty-two and eighty-four germinated seedlings showed albino and wild type phenotype, respectively. Chi-square analysis ($\chi^2 = 0.414$, $p = 0.520$) indicated that the segregation ratio between albino and wild type was 1:3. Therefore, the albino trait is likely controlled by a single recessive nuclear gene.

Transcriptome profiling and identifying differentially expressed genes (DEGs) between the *alc* mutant and wild type seedlings

The transcriptomes of cotyledons from the *alc* mutant and wild type seedlings were examined by RNA-seq, each with three biological replicates. Overall, 97,869,948 to 123,516,412 clean reads were obtained after filtering low quality reads. After mapping to the cucumber reference genome 9930 v2 (Huang et al., 2009; Li et al., 2011), 21,664 transcripts were identified. High correlation coefficients among the replicates demonstrated the consistency of the transcriptional changes within each sample (Figure 3A). In total, 1,256 genes were upregulated and 1,584 were downregulated in the *alc* mutant compared to the wild type cotyledons ($|\log2FC| \geq 2$)

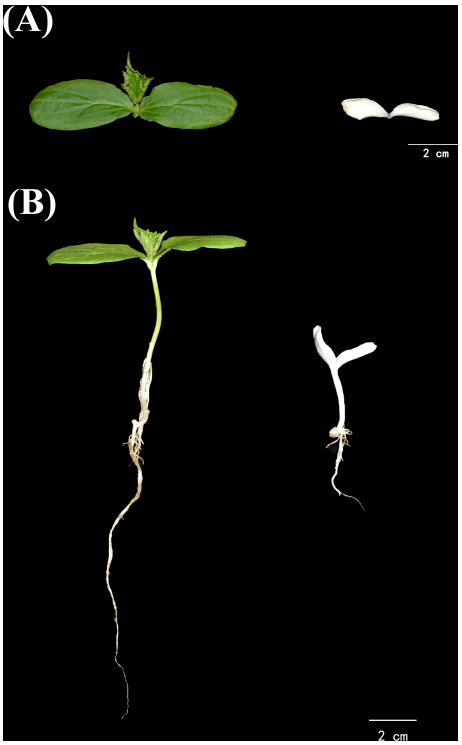


FIGURE 1
Growth phenotype of *alc* and wild type from the progenies of a cucumber inbred line g32. **(A)** Different cotyledons of *alc* and wild type seedlings. **(B)** Morphological difference between *alc* and wild type seedlings.

(Figure 3B; Supplementary Table 2). Based on the annotation, DEGs were annotated using GO and KEGG pathway to identify the significantly enriched biological processes and pathways between the *alc* mutant and wild type. In total, 2,175 DEGs were classified into 814 GO terms belonging to three categories: biological process, cellular component, and molecular function (Supplementary Table 3). Cellular carbohydrate biosynthetic process (GO:0034637, p=0.00010) and cellular carbohydrate metabolic process (GO:0044262, p=0.00017) were the most significantly enriched biological processes (Supplementary Figure 1). As shown in Figure 4, 171 and 196 DEGs were identified as transmembrane transport (GO:0055085) and transporter activity (GO:0005215), respectively, which were among the most enriched GO terms. (Supplementary Table 3). A total of, 957 DEGs were assigned to 110 KEGG pathways (Supplementary Table 4), among which, top 20 enriched KEGG pathways are illustrated in Figure 4. Carbon metabolism (KEGG: csv01200), phenylpropanoid biosynthesis (KEGG: csv00940), and starch and sucrose metabolism (KEGG: csv00500) pathways contained 69, 50, and 44 DEGs, respectively (Supplementary Table 4).

Differentially expressed genes involved in chlorophyll metabolism

Porphyrin and chlorophyll metabolism (KEGG: csv00860) was a significantly enriched pathway in KEGG analysis. Many key enzymes involved in this pathway showed distinct expression profile between the *alc* mutant and the wild type. The expression of most genes, including *HEMB* (Csa2G401270), *HEME* (Csa4G082410, Csa5G218840), *HEMF* (Csa4G056670), *HEMG* (Csa6G007980), CHLG (Csa4G311220), and CAO (Csa6G385090), was slightly higher in the *alc* mutant than in the wild type (Figure 5A; Supplementary Table 5). However, POR (Csa4G638340) was downregulated in the *alc* mutant, presenting an opposite expression pattern to that of other genes (Figure 5A; Supplementary Table 5).

Differentially expressed genes involved in methylerythritol 4-phosphate (MEP) pathway

The methylerythritol 4-phosphate (MEP) pathway is mainly involved in the production of isoprenoid precursors; isopentenyl diphosphate (IPP) and dimethylallyl diphosphate in photosynthetic eukaryotes (Cordoba et al., 2009). As shown in Figure 5B, there were at least seven key enzymes involved in the MEP pathway. Four candidate genes encoding CMS (Csa3G113320 and Csa4G049620), CMK (Csa1G600780), and MCS (Csa4G049620) were downregulated in the *alc* mutant (Supplementary Table 5).

Chloroplast development-related genes

Chloroplast-localized FtsH proteins are crucial for the biogenesis of thylakoid membranes (Wagner et al., 2012). However, we observed negligible expression of the two *FtsH genes* (Csa6G504470 and Csa6G504480) in the *alc* mutant (Supplementary Table 2). Some pentatricopeptide repeat (PPR) proteins are involved in plastid gene expression and can also affect chloroplast development (Myouga et al., 2010). The expression of *PPR* gene (Csa5G189930) in the *alc* mutant was more than 3.5-fold higher than that in the wildtype (Supplementary Table 2).

Thylakoid related genes were affected in *alc* mutant

Nineteen DEGs involved in thylakoid related functional activities, including thylakoid (GO:0009579), thylakoid part (GO:0044436), and thylakoid membrane (GO:0042651), were highlighted (Figure 6; Supplementary Table 3). Three *PsbPs*

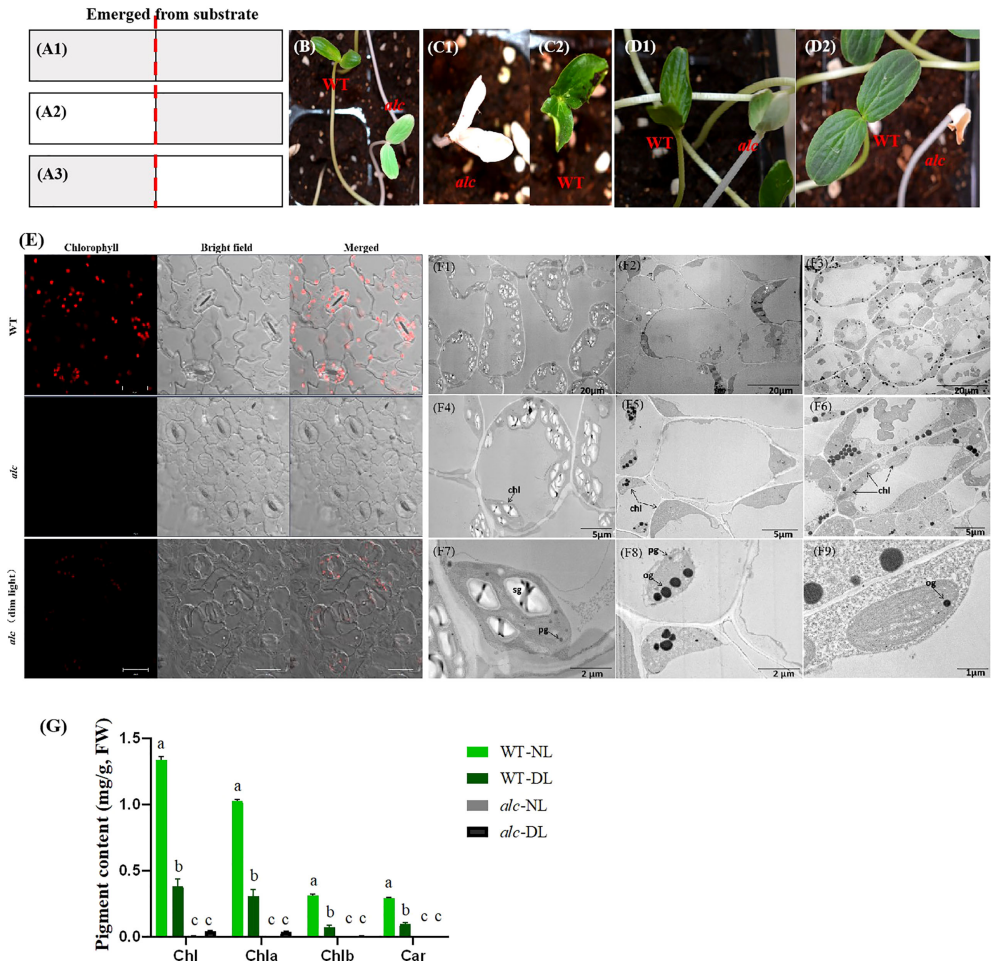


FIGURE 2
Short-lived chloroplast recovery in *alc* mutant under dim light condition. **(A1–A3)** Schematic illustration of light treatments of WT and *alc* seedlings. A1 Plants were treated with continuous dim light before and after they emerged from substrate. A2 First, plants were treated with normal light irradiation, then after emerging from substrate, they were moved to dim light condition. A3 First, plants were treated under dim light, then after emerging from substrate, they were moved to normal light condition. **(B)** Phenotype of WT and *alc* seedlings under indicated condition that described in a1. **(C)** Phenotype of WT **(C1)** and *alc* **(C2)** seedlings under indicated conditions that described in A2. **(D)** Phenotype of WT and *alc* seedlings under indicated conditions that described in A3. D1 Phenotype of WT and *alc* seedlings exposed to normal light after 5 hours. D2 Phenotype of WT and *alc* seedlings exposed to normal light after 30 hours. **(E)** Fluorescence microscopy images of cotyledon abaxial epidermis of wild type, *alc* mutant cucumber seedlings grown under normal light condition and *alc* mutant seedlings grown under dim light condition. Bar: 20μm. **(F1–F9)** Transmission electron microscopy of cotyledons from *alc* mutant and wild type seedlings. F1, F4 An overview of cotyledon cells of wild type grown under normal light condition. **(F2–F5)** An overview of cotyledon cells of *alc* mutant grown under normal light condition. F3, F6 An overview of cotyledon cells of *alc* mutant grown under dim light condition. **(F7–F9)** Enlarged views of chloroplast ultrastructure of wild type, *alc* mutant grown under normal light condition and *alc* mutant grown under dim light condition, respectively. chl, chloroplast sg, starch granules pg, plastoglobuli og, osmiophilic plastoglobuli. **(G)** Chlorophyll content and carotenoid content of wild type and *alc* mutant grown under normal light and dim light conditions. WT-NL, wild type grown under normal light condition. WT-DL, wild type grown under dim light condition alc-NL, *alc* mutant grown under normal light condition alc-DL, *alc* mutant grown under dim light condition. Chla, chlorophyll a; Chlb, chlorophyll b; Chl, total chlorophyll.

genes (Csa2G030040, Csa1G088470, and Csa1G181310) were upregulated in the *alc* mutant, while 16 other genes, namely, *PsbR* (Csa4G064020), *PsaG/PsaK* (Csa3G060980, Csa6G525340), *PsbO* (Csa6G488340), *PsaD* (Csa3G147780), *PsaH* (Csa3G483830), *PsbP* (Csa4G063440), *PsaE* (Csa2G079660), *PsbY* (Csa5G592810), *PsbW* (Csa7G378440), *PsaN* (Csa6G483300), *PsbQ* (Csa1G066480, Csa3G414060), *PsbX* (Csa1G595840), *PsaF* (Csa1G714680), and *PetM* (Csa7G075020) were downregulated in the *alc* mutant (Supplementary Table 2).

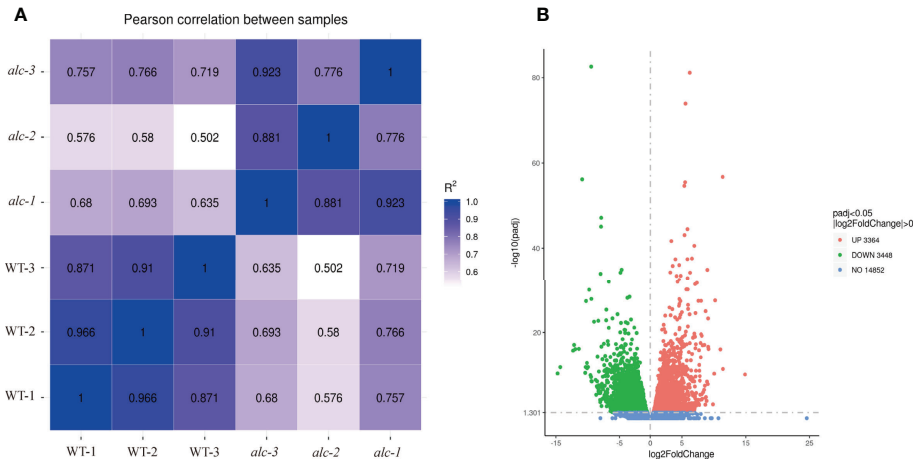


FIGURE 3
Diagrams illustrating correlations/distances among transcriptomes and the number of differentially expressed genes in *alc* mutant and wild type seedlings. **(A)** Correlation matrix and cluster dendrogram of the whole dataset of the mapped reads. The analysis was performed by comparing the values of the entire transcriptome of all two samples with three biological replicates. Correlation analysis (coefficients R^2) and hierarchical cluster analysis were performed using R software. Dark blue color indicated a stronger correlation and light blue weaker (R^2). **(B)** Volcano plot showing DEGs between *alc* mutant and wild type. X-axis represented log2(Fold Change) and y-axis represents -log10 (padj). Red, green and blue dots represented up-regulated, down-regulated and not DEGs, respectively.

Homologous genes from albino mutants in other plant species

Previous studies have reported that mutations of FRUCTOKINASE-LIKE PROTEIN in barley and rice (Qin et al., 2015; Lv et al., 2017; He et al., 2018), *RPL21c* (chloroplast 50S ribosomal protein L21) in Arabidopsis and rice (Yin et al., 2012; Lin et al., 2015), *EMB* (embryo-defective) in Arabidopsis (Huang et al., 2009; Liang et al., 2010; Ye et al., 2017; Chen et al., 2018), *PDS3* (phytoene desaturase) in Arabidopsis (Qin et al., 2007), *TOC159* (Translocase of chloroplast 159) in Arabidopsis (Kakizaki et al., 2009; Shanmugabalaji et al., 2018), *DXS1* (1-deoxy-D-xylulose-5-phosphate synthase 1) in tomato (Garcia-Alcazar et al., 2017), and *PurD* (phosphoribosylamine–glycine ligase) in rice (Zhang et al., 2018) can cause albinism. The coding sequences (CDSs) of the above mentioned homologous genes of *alc* mutant and wild type were compared

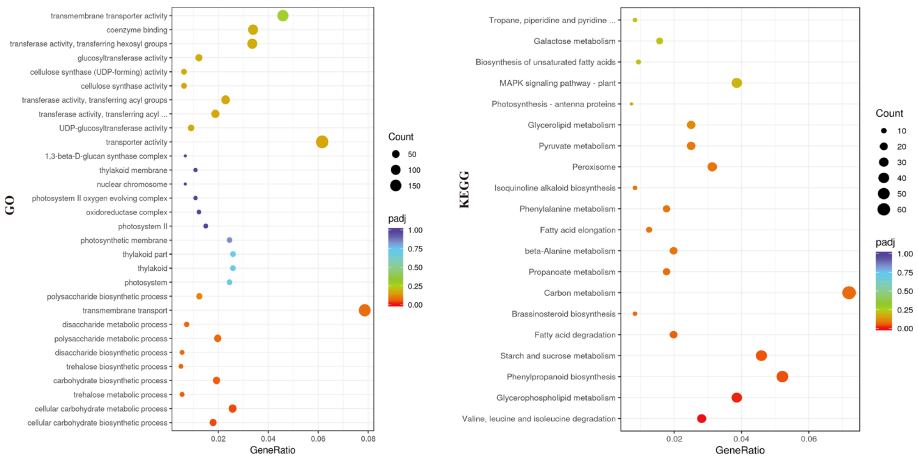


FIGURE 4
Genes enriched in different categories in the GO (left) and KEGG (right) analysis. X-axis represented the gene ratio of enriched genes among the background genes in different categories and y-axis represented the GO or KEGG terms. A high padj-value was represented by red, and a low value represented by purple. The size of the bubble represented number of genes annotated to each term.

using our transcriptome data. However, no variant was found among these genes in the*alc* mutant and wildtype. Among these genes, only *TOC159* (Csa4G001670) and *DXS1* (Csa3G114510) showed more abundant expression in *alc* mutant than in the wild-type (Supplementary Table 6).

Validation of DEG expression by RT-qPCR

Twelve genes including seven genes involved in chlorophyll metabolism, three genes from the MEP pathway, *TOC159*, and *DXS1* were selected for RT-qPCR verification (the information of the genes is listed in Supplementary Table 1). The relative expression (−ΔΔCt) of each gene was calculated using corresponding wild type gene expressions as control. The correlation between the relative expression value and RNA-seq result (log2FoldChange) of the *alc* mutant was determined. The correlation between RT-qPCR and RNA-seq data was 0.9780 ($P < 0.0001$, ****), indicating the reliability of our RNA-seq results.

Discussion

Albinism occurs among different living organisms ranging from human beings to animals as well as higher plants (Kamaraj and Purohit, 2014; Galante Rocha de Vasconcelos et al., 2017; Le

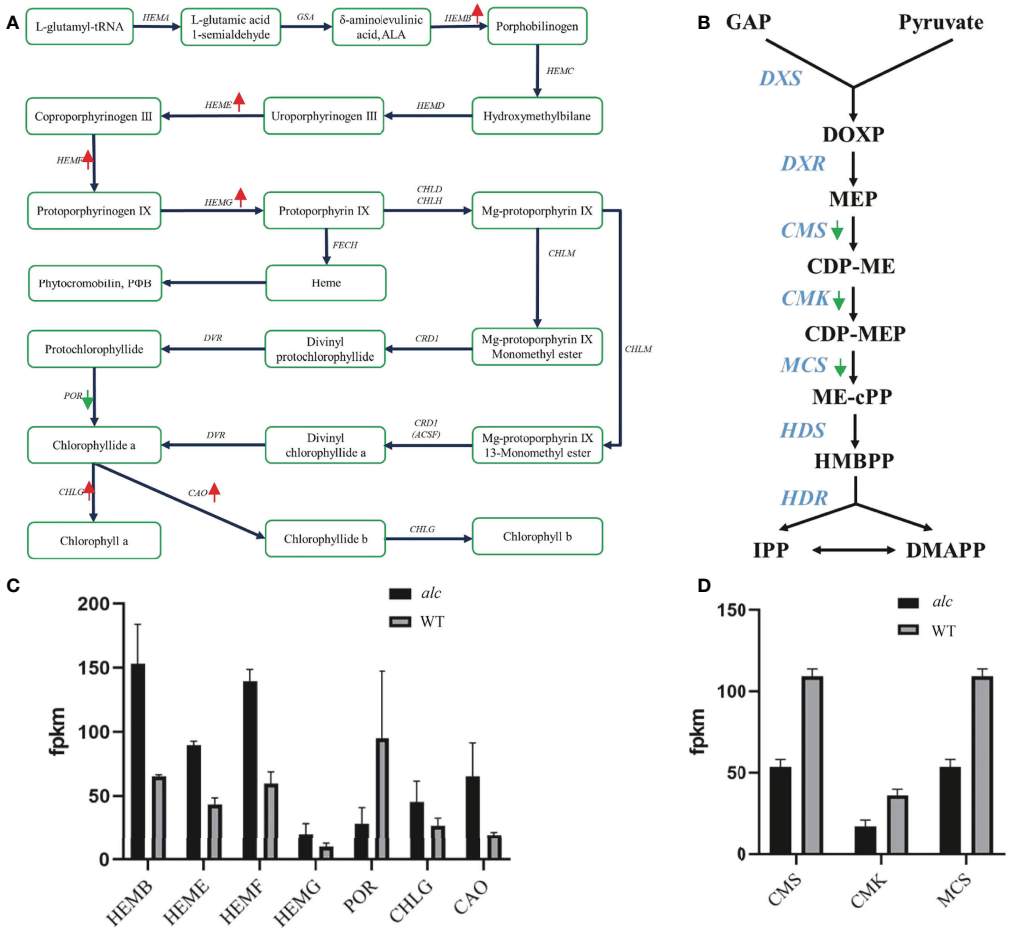


FIGURE 5
Differential expression of genes involved in chlorophyll metabolism and Methylerythritol 4-phosphate (MEP) pathway. **(A, B)** Diagram showing major genes in chlorophyll metabolism and MEP pathway, respectively. Red arrows indicated up-regulated genes and green arrows indicated down-regulated genes in *alc* mutant. **(C, D)** Expression profile of DEGs in chlorophyll metabolism and MEP pathway achieved by RNA-seq. The mean FPKM (fragments per kilo base of transcript per million mapped reads) values for the DEGs were calculated from three biological replicates for each genotype. Error bars indicated standard deviations. GAP, glyceraldehyde-3-phosphate; DOXP, 1-deoxy-D-xylulose-5-phosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate; CDP-ME, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate; ME-cPP, 2-C-methyl-D-erythritol-2,4-C-cyclodiphosphate; HMBPP, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate; IPP, isopentenyl diphosphate DMAPP, dimethylallyl diphosphate.

Ret et al., 2018). In the present study, some albino seedlings were observed in the progenies of a self-pollinated cucumber fruit during seed increase period of a cucumber inbred line named "g32". The mutant was named *alc*, which presented white cotyledons and hypocotyl, and died before developing first true leaf under normal light conditions. In most studies, albino plants were caused by lack of chlorophyll and impaired chloroplast development (Liang et al., 2010; Zhu et al., 2015; Le Ret et al., 2018). Consistent with most albino mutants, lack of chlorophyll and defective chloroplast development was also observed in the *alc* mutant (Figure 2). In wild-type seedlings, ultrastructure of chloroplast was well presented with compactly arranged chloroplasts, while few or even no chloroplasts were observed in albino seedlings. Furthermore, an abnormal chloroplast ultrastructure lacking starch granules and thylakoids, but with osmiophilic plastoglobuli was observed in the *alc* mutant (Figure 4). Osmiophilic plastoglobuli generally appear as a result of the degradation of thylakoid membranes under stress (Wang et al., 2016) Therefore, we propose that irradiation with normal light might act as an abiotic stress cue for *alc* mutant, leading to the degradation of thylakoid membranes.

Interestingly, we found that the cucumber *alc* mutant differed from the reported *fln1, rpl21c, emb, pds3, toc159, dxs1, and purd* albino mutants as this mutant presented green cotyledons under dim light conditions. Additionally, we checked the homologous gene sequence of these albino related genes in cucumber, but found no variation within the coding sequences of the *alc* mutant and wild type, suggesting that *alc* might be a novel light sensitive albino mutant. Similar phenotype was reported in *pap7-1* albino mutant of Arabidopsis (Grübler et al., 2017). *Pap7-1* mutant had an albino cotyledon and died grown under light conditions; however, under dark conditions, *pap7-1* could grow when supplied with sucrose supplemented medium (Grübler et al., 2017). However, the molecular mechanism describing the exact nature of *pap7-1* is still unknown.

Since all the seeds were planted under the same culture condition, we could exclude the possible of environmental influences on the albino phenotype of the *alc* mutant. Genetic analysis revealed that this albino phenotype was controlled by a recessive locus. Phenotypic evaluation of the previous generation of "g32" was also performed but albino seedlings were not observed (data not shown), confirming the recessive inheritance of this albino allele. Albinism is not a desired phenomenon in plant breeding since it could affect plant growth as well as production. However, this *alc* mutant is of great importance to us for detection of new genes controlling plastid development.

Transcriptome analysis has been extensively applied to identify major genes and dissect regulatory networks involved in albinism by using different albino mutants (Shi et al., 2017; Li et al., 2017; Ma et al., 2018). In our study, transcriptome analysis revealed that the expression of genes involved in chloroplast development, chlorophyll metabolism, MEP pathway, and other genes, such as glutathione S-transferase, was altered in the *alc* mutant. This transcriptional modification suggests that these genes play important roles for the albino phenotype of the *alc* mutant.

Light plays a crucial role in plant development. In light-sensitive mutants, a sudden increase in ROS production would occur under excessive light and induce oxidative damage leading to leaf bleaching. Arabidopsis *ch1* is a Chlb deficient mutant and is devoid of photosystem II (PSII) Chl-protein antenna complexes. Thus, oxidizing side of PSII is impaired and *ch1* is sensitive to photooxidative stress (Havaux et al., 2007). Nearly all DEGs in the Photosynthesis-antenna proteins pathway (Supplementary Table 4, KEGG: csv00196) were downregulated in the *alc* mutant, indicating a weak functionality of core reaction center. Peroxisomes are organelles that contribute to the reduction of oxidative stress (del Río et al., 2006). Most of genes in Peroxisome pathway (24 out of 30) (Supplementary Table 4, KEGG: csv04146) were upregulated in the *alc* mutant. Under normal light irradiation for wild-type but excessive for *alc*, the *alc* might produce more ROS and lead to bleaching and finally died as *ch1* (Ramel et al., 2013). Since *pap7-1* mutant could be arrested under very dim light (Grübler et al., 2017), we tried to culture the mutant under dim light condition but failed. Further studies are needed to confirm whether the morphology and flowering of the mutant are similar to that of *pap7-1* grown under sucrose supplemented medium and dim light conditions.

Three genes involved in chloroplast development were found to show different transcript levels between the *alc* mutant and wild type. Two of these genes belong to FtsH family, which is essential for chloroplast development. For example, both *ftsh1 ftsf5* and *ftsh2 ftsh8* double mutants developed white seedlings with disrupted chloroplast development (Zaltsman et al., 2005). Significantly fewer *FstH* transcripts were detected in the *alc* mutant as compared to those in wild-type cucumber. E*ntratricopeptide Repeat Protein Pigment-Defective Mutant2 (PDM2)*, that encodes a PPR protein, is required for chloroplast development by regulation of plastid gene expression (Du et al., 2017). We detected very low expression of a *PPR* gene in the *alc* mutant. The extremely low expression of chloroplast related genes might impact chloroplast development and result in an albino phenotype. Key genes involved in chlorophyll metabolism, the MEP pathway, and thylakoid function were also affected in the *alc* mutant. Many genes in the chlorophyll metabolism pathway, including *HEMB*, *HEME*, *HEMF*, *HEMG*, and *CHLG* were slightly up regulated in the *alc* mutant. The increased expression

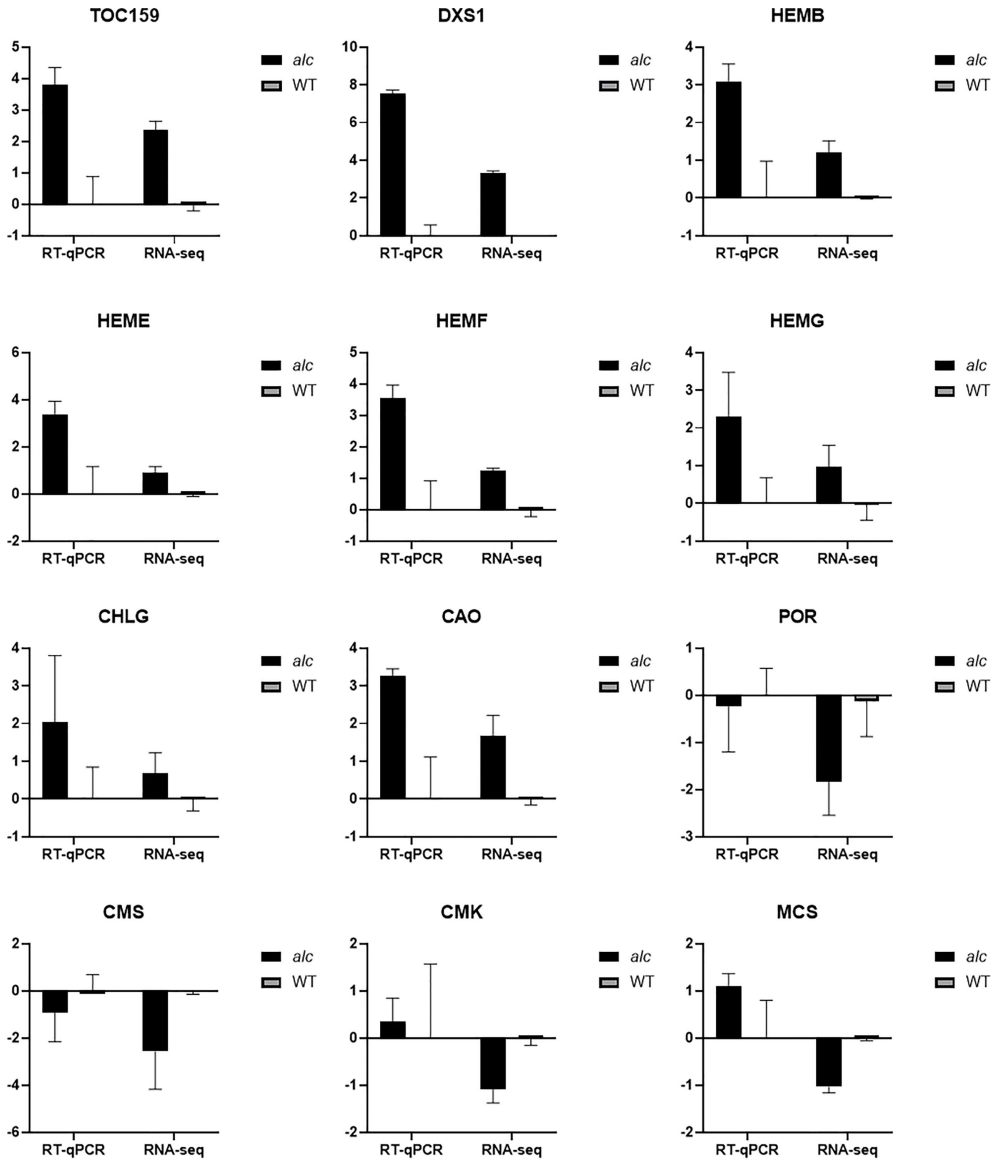


FIGURE 6
Expression profile of selected differentially expressed genes between *alc* mutant and wild type from RNA-seq result achieved by qRT-PCR. X-axis represented gene name and y-axis represented relative expression (−ΔΔCt) value of each gene. Data are shown as means (n = 3). Error bars indicated standard deviations.

of these genes was also observed in other albino mutants. Most of the chlorophyll-biosynthesis related genes, such as *HEMC*, *HEME* and *CHLG* were upregulated in white leaves compared with those in green leaves of *Ananas comosus* var. *Bracteatus* (Li et al., 2017). Similar results were observed in the wheat *mta* albino mutant where the expression of *HEME* was upregulated (Shi et al., 2017). The decreased expression of most genes might be regulated by a feedback mechanism. *POR*, a key light-dependent enzyme, is essential to chlorophyll biosynthesis where it catalyzes protochlorophyllide to chlorophyllide (Yang and Cheng, 2004; Zhang et al., 2019). POR is crucial for plant growth and development because *por* mutant and POR RNAi line displayed reduced chlorophyll content and severe photoautotrophic growth defects (Kim and Apel, 2012; Paddock et al., 2012). The expression level of POR was low in the *alc* mutant, the *mta* albino wheat mutant (Shi et al., 2017), and the complete white leaves of *Ananas comosus* var. *Bracteatus* (Li et al., 2017), indicating that the downregulation of POR might have an impact on chlorophyll synthesis in the *alc* mutant. The MEP pathway is essential for the biosynthesis of photosynthesis-related compounds, such as

carotenoids, chlorophylls, gibberellins, and abscisic acid, which are of vital importance for plant development and metabolism (Flores-Perez et al., 2008). The mutation of genes in the pathway impaired the biosynthesis of these compounds, disrupted chloroplast development, and resulted in abnormal plant morphology, especially in leaf color (Xing et al., 2010; Chandran et al., 2016). IspD (CMS), IspE (CMK), and IspF (MCS) are the third, fourth and fifth enzymes in the MEP pathway, respectively (Figure 5B). Related mutants possessed yellow or albino leaves with arrested development of chloroplasts (Hsieh et al., 2008; Chen et al., 2018; Huang et al., 2018). In this study, the expression of CMS, CMK and MCS in the *alc* mutant was lower than that in the wildtype. The low-level expression may contribute to the disruption of chloroplast development. Thylakoid related genes corresponded to reaction centers of PSI and PSII where photochemical reactions occur and convert light energy into chemical energy. Almost all genes related to thylakoid were downregulated in the *alc* mutant, demonstrating the decrease of photosynthesis viability.

Isocitrate lyase and malate synthase are two enzymes unique to the glyoxylate cycle, which is considered essential for postgerminative growth and seedling establishment (Eastmond et al., 2000). Under normal light conditions, the enzyme activities in the glyoxylate cycle decreased rapidly as seedlings become photosynthetic, whereas those under dark conditions maintained a continuously high level of enzyme activity (Trelease et al., 1971; Becker et al., 1978; He et al., 2016). In this study, the abundance of isocitrate lyase (Csa2G420990) and malate synthase (Csa1G050360) was higher in the mutant than in the wildtype (Figure 6). The wild-type exhibited autotrophic growth with a high activity of peroxidase (Csa4G285740) and chlorophyll A-B binding protein (Csa3G664560), while the *alc* mutant was unable to perform photosynthesis due to the lack of functional chloroplasts. Glutathione S-transferase have been reported to mainly function in response to biotic and abiotic stresses, such as oxidative stress (Horváth et al., 2015), temperature stress (Roxas et al., 1997; Liang et al., 2018) and different pathogen invasion (Sytykiewicz et al., 2014; Gong et al., 2018; Ducker et al., 2019). The high concentration of glutathione S-transferase can result in decreased chlorophyll content (Jiang et al., 2010). When the *alc* mutant is exposed to light, which is lethal rather than beneficial, it expressed (Csa4G304240, Csa4G303170) (Figure 6) more glutathione S-transferase than the wild type; this might also promote chlorophyll degradation.

In conclusion, a cucumber albino mutant *alc* was cytologically, genetically and transcriptomically characterized in this study. Our results demonstrated that the albino phenotype of the mutant was mainly due to the disability in chlorophyll synthesis and chloroplast development, which resulted in no chlorophyll content in the cotyledons and finally seedlings died as not being able to photosynthesize. In this study, we could not determine the causal mutation of *alc*. However, Bulk Segregant Analysis (BSA), using albino and green seedling pools along with map-based cloning would be useful to finely mapping the mutated gene in future studies.

Data availability statement

The RNA-seq datasets presented in this study can be accessed through National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA685868.

Author contributions

JY, BL, YL, and BJ: conceptualization and writing—review and editing; JY, ZC, LC, ZL, MW, and WL: formal analysis and investigation; JY: writing—original draft preparation. All authors have read and approved the final manuscript.

Funding

This study was supported by Key-Area Research and Development Program of Guangdong Province (2020B020220001), The Discipline Team Construction Project of GDAAS (202103TD), the Training Plan for Young and Middle-aged Discipline Leaders of GDAAS (R2020PY-JG003) and Talent Introduction Plan of GDAAS (R2021YJ-YB2004). The funding bodies have no role in the study design, data analysis and interpretation, and manuscript writing, but just provide the financial supports.

Acknowledgments

The authors thank associate professor Yuhui Wang from College of Horticulture at Nanjing Agricultural University for her assistance in 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

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.2022.1047090/full#supplementary-material

SUPPLEMENTARY TABLE 1
Primers used for qRT-PCR validation.

SUPPLEMENTARY TABLE 2
Up-regualted and down-regualted genes in albino mutant.

SUPPLEMENTARY TABLE 3
GO enrichment analysis of differentially expressed genes between albino mutant and wild type.

SUPPLEMENTARY TABLE 4
KEGG analysis of differentially expressed genes between albino mutant and wild type.

SUPPLEMENTARY TABLE 5
Differentially expressed genes in chlorophyll metabolism and MEP pathway between albino mutant and wild type.

SUPPLEMENTARY TABLE 6
Differentially expressed genes of reported albino genes between albino mutant and wild type.

SUPPLEMENTARY FIGURE 1
Histogram of GO enrichment analysis.

SUPPLEMENTARY FIGURE 2
A heat map showing expression patterns of genes related to chlorophyll metabolism, chloroplast formation and Methylerythritol 4-phosphate (MEP) pathway.

References

Becker, W. M., Leaver, C. J., Weir, E. M., and Riezman, H. (1978). Regulation of glyoxysomal enzymes during germination of cucumber: I. developmental changes in cotyledonary protein, RNA, and enzyme activities during germination. *Plant Physiol.* 62 (4), 542–549. doi: 10.1104/pp.62.4.542

Chandran, A. K. N., Lee, G. S., Yoo, Y. H., Yoon, U. H., Ahn, B. O., Yun, D. W., et al. (2016). Functional classification of rice flanking sequence tagged genes using MapMan terms and global understanding on metabolic and regulatory pathways affected by *dxr* mutant having defects in light response. *Rice (N Y)* 9 (1), 17. doi: 10.1186/s12284-016-0089-2

Chan, K. X., Phua, S. Y., Crisp, P., McQuinn, R., and Pogson, B. J. (2016). Learning the languages of the chloroplast: Retrograde signaling and beyond. *Annu. Rev. Plant Biol.* 67, 25–53. doi: 10.1146/annurev-arplant-043015-111854

Chen, H., Li, S., Li, L., Hu, H., and Zhao, J. (2018). Arabidopsis *EMB1990* encoding a plastid-targeted YlmG protein is required for chloroplast biogenesis and embryo development. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00181

Chen, N., Wang, P., Li, C., Wang, Q., Pan, J., Xiao, F., et al. (2018). A single nucleotide mutation of the *IspE* gene participating in the MEP pathway for isoprenoid biosynthesis causes a green-revertible yellow leaf phenotype in rice. *Plant Cell Physiol.* 59 (9), 1905–1917. doi: 10.1093/pcp/pcy108

Cordoba, E., Salmi, M., and Leon, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60 (10), 2933–2943. doi: 10.1093/jxb/erp190

del Río, L. A., Corpas, F. J., Sandalio, L. M., Palma, J. M., Gómez, M., and Barroso, J. B. (2002). Reactive oxygen species, antioxidant systems and nitric oxide in peroxisomes. *J. Exp. Bot.* 53 (372), 1255–1272. doi: 10.1093/jexbot/53.372.1255

Ducker, R., Zollner, P., Lummen, P., Ries, S., Collavo, A., and Beffa, R. (2019). Glutathione transferase plays a major role in flufenacet resistance of ryegrass (Lolium spp.) field populations. *Pest Manag Sci.* 75 (11), 3084–3092. doi: 10.1002/ps.5425

Du, L., Zhang, J., Qu, S., Zhao, Y., Su, B., Lv, X., et al. (2017). The *Pentratricopeptide repeat protein pigment-defective Mutant2* is involved in the regulation of chloroplast development and chloroplast gene expression in arabidopsis. *Plant Cell Physiol.* 58 (4), 747–759. doi: 10.1093/pcp/pcx004

Eastmond, P. J., Germain, V., Lange, P. R., Bryce, J. H., Smith, S. M., and Graham, I. A. (2000). Postgerminative growth and lipid catabolism in oilseeds lacking the glyoxylate cycle. *Proc. Natl. Acad. Sci. U.S.A.* 97 (10), 5669–5674. doi: 10.1073/pnas.97.10.5669

Flores-Perez, U., Sauret-Gueto, S., Gas, E., Jarvis, P., and Rodriguez-Concepcion, M. (2008). A mutant impaired in the production of plastome-encoded proteins uncovers a mechanism for the homeostasis of isoprenoid biosynthetic enzymes in arabidopsis plastids. *Plant Cell* 20 (5), 1303–1315. doi: 10.1105/tpc.108.058768

Galante Rocha de Vasconcelos, F. T., Hauzman, E., Dutra Henriques, L., Kilpp Goulart, P. R., de Faria Galvao, O., Sano, R. Y., et al. (2017). A novel nonsense mutation in the tyrosinase gene is related to the albinism in a capuchin monkey (Sapajus apella). *BMC Genet.* 18 (1), 39. doi: 10.1186/s12863-017-0504-8

Garcia-Alcazar, M., Gimenez, E., Pineda, B., Capel, C., Garcia-Sogo, B., Sanchez, S., et al. (2017). Albino T-DNA tomato mutant reveals a key function of 1-deoxy-D-xylulose-5-phosphate synthase (DXS1) in plant development and survival. *Sci. Rep.* 7, 45333. doi: 10.1038/srep45333

Gong, Q., Yang, Z., Chen, E., Sun, G., He, S., Butt, H. I., et al. (2018). A phi-class glutathione s-transferase gene for verticillium wilt resistance in gossypium arboreum identified in a genome-wide association study. *Plant Cell Physiol.* 59 (2), 275–289. doi: 10.1093/pcp/pcx180

Gotoh, E., Suetsugu, N., Yamori, W., Ishishita, K., Kiyabu, R., Fukuda, M., et al. (2018). Chloroplast accumulation response enhances leaf photosynthesis and plant biomass production. *Plant Physiol.* 178 (3), 1358–1369. doi: 10.1104/pp.18.00484

Grübler, B., Merendino, L., Twardziok, S. O., Mininno, M., Allorent, G., Chevalier, F., et al. (2017). Light and plastid signals regulate different sets of genes in the albino mutant *Pap7-1*. *Plant Physiol.* 175 (3), 1203–1219. doi: 10.1104/pp.17.00982

Havaux, M., Dall'osto, L., and Bassi, R. (2007). Zeaxanthin has enhanced antioxidant capacity with respect to all other xanthophylls in arabidopsis leaves and functions independent of binding to PSII antennae. *Plant Physiol.* 145 (4), 1506–1520. doi: 10.1104/pp.107.108480

He, D., Damaris, R. N., Fu, J., Tu, J., Fu, T., Xi, C., et al. (2016). Differential molecular responses of rapeseed cotyledons to light and dark reveal metabolic adaptations toward autotrophy establishment. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2016.00988

He, L., Zhang, S., Qiu, Z., Zhao, J., Nie, W., Lin, H., et al. (2018). FRUCTOKINASE-LIKE PROTEIN 1 interacts with TRXz to regulate chloroplast development in rice. *J. Integr. Plant Biol.* 60 (2), 94–111. doi: 10.1111/jipb.12631

Horváth, E., Bela, K., Papdi, C., Gallé, Á., Szabados, L., Tari, I., et al. (2015). The role of arabidopsis glutathione transferase F9 gene under oxidative stress in seedlings. *Acta Biol. Hung* 66 (4), 406–418. doi: 10.1556/018.66.2015.4.5

Hsieh, M. H., Chang, C. Y., Hsu, S. J., and Chen, J. J. (2008). Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in *ispD* and *ispE* albino mutants in arabidopsis. *Plant Mol. Biol.* 66 (6), 663–673. doi: 10.1007/s11103-008-9297-5

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., et al. (2009). The genome of the cucumber, cucumis sativus l. *Nat. Genet.* 41 (12), 1275–1281. doi: 10.1038/ng.475

Huang, R., Wang, Y., Wang, P., Li, C., Xiao, F., Chen, N., et al. (2018). A single nucleotide mutation of *IspF* gene involved in the MEP pathway for isoprenoid biosynthesis causes yellow-green leaf phenotype in rice. *Plant Mol. Biol.* 96 (1-2), 5–16. doi: 10.1007/s11103-017-0668-7

Huang, X., Zhang, X., and Yang, S. (2009). A novel chloroplast-localized protein EMB1303 is required for chloroplast development in arabidopsis. *Cell Res.* 19 (10), 1205–1216. doi: 10.1038/cr.2009.84

Iida, S., and Amano, E. (1991). Mutants induced by pollen irradiation in cucumber. *Cucurbit Genet. Coop. Rpt* 14, 32–33.

Jiang, H. W., Liu, M. J., Chen, I. C., Huang, C. H., Chao, L. Y., and Hsieh, H. L. (2010). A glutathione s-transferase regulated by light and hormones participates in the modulation of arabidopsis seedling development. *Plant Physiol.* 154 (4), 1646–1658. doi: 10.1104/pp.110.159152

Kakizaki, T., Matsumura, H., Nakayama, K., Che, F. S., Terauchi, R., and Inaba, T. (2009). Coordination of plastid protein import and nuclear gene expression by plastid-to-nucleus retrograde signaling. *Plant Physiol.* 151 (3), 1339–1353. doi: 10.1104/pp.109.145987

Kamaraj, B., and Purohit, R. (2014). Mutational analysis of oculocutaneous albinism: A compact review. *BioMed. Res. Int.* 2014, 905472. doi: 10.1155/2014/905472

Kim, C., and Apel, K. (2012). Arabidopsis light-dependent NADPH: protochlorophyllide oxidoreductase a (PORA) is essential for normal plant growth and development: an addendum. *Plant Mol. Biol.* 80 (2), 237–240. doi: 10.1007/s11103-012-9944-8

Le Ret, M., Belcher, S., Graindorge, S., Wallet, C., Koechler, S., Erhardt, M., et al. (2018). Efficient replication of the plastid genome requires an organellar thymidine kinase. *Plant Physiol.* 178 (4), 1643–1656. doi: 10.1104/pp.18.00976

Liang, D., Gao, F., Ni, Z., Lin, L., Deng, Q., Tang, Y., et al. (2018). Melatonin improves heat tolerance in kiwifruit seedlings through promoting antioxidant enzymatic activity and glutathione s-transferase transcription. *Molecules* 23 (3), 584. doi: 10.3390/molecules23030584

Liang, Q., Lu, X., Jiang, L., Wang, C., Fan, Y., and Zhang, C. (2010). EMB1211 is required for normal embryo development and influences chloroplast biogenesis in arabidopsis. *Physiol. Plant* 140 (4), 380–394. doi: 10.1111/j.1399-3054.2010.01407.x

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: pigment of photosynthetic biomembranes. *Method Enzymol.* 148, 350–382. doi: 10.1016/0076-6879(87)48036-1

Li, X., Kanakala, S., He, Y., Zhong, X., Yu, S., Li, R., et al. (2017). Physiological characterization and comparative transcriptome analysis of white and green leaves of *Ananas comosus* var. bracteatus. *PloS One* 12 (1), e0169838. doi: 10.1371/journal.pone.0169838

Li, Z., Mo, W., Jia, L., Xu, Y. C., Tang, W., Yang, W., et al. (2019). Rice FLUORESCENT1 is involved in the regulation of chlorophyll. *Plant Cell Physiol.* 60 (10), 2307–2318. doi: 10.1093/pcp/pcz129

Lin, D., Jiang, Q., Zheng, K., Chen, S., Zhou, H., Gong, X., et al. (2015). Mutation of the rice *ASL2* gene encoding plastid ribosomal protein L21 causes chloroplast developmental defects and seedling death. *Plant Biol. (Stuttg)* 17 (3), 599–607. doi: 10.1111/plb.12271

Liu, B., Liu, X., Yang, S., Chen, C., Xue, S., Cai, Y., et al. (2016). Silencing of the gibberellin receptor homolog, CsGID1a, affects locule formation in cucumber (Cucumis sativus) fruit. *New Phytol.* 210 (2), 551–563. doi: 10.1111/nph.13801

Li, Z., Zhang, Z., Yan, P., Huang, S., Fei, Z., and Lin, K. (2011). RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics* 12, 540. doi: 10.1186/1471-2164-12-540

Lu, M., Han, J., Zhu, B., Jia, H., Yang, T., Wang, R., et al. (2019). Significantly increased amino acid accumulation in a novel albino branch of the tea plant (*Camellia sinensis*). *Planta* 249 (2), 363–376. doi: 10.1007/s00425-018-3007-6

Lv, Y., Shao, G., Qiu, J., Jiao, G., Sheng, Z., Xie, L., et al. (2017). White leaf and panicle 2, encoding a PEP-associated protein, is required for chloroplast biogenesis under heat stress in rice. *J. Exp. Bot.* 68 (18), 5147–5160. doi: 10.1093/jxb/erx332

Ma, Q., Li, H., Zou, Z., Arkorful, E., Lv, Q., Zhou, Q., et al. (2018). Transcriptomic analyses identify albino-associated genes of a novel albino tea germplasm 'Huabai 1'. *Hortic. Res.* 5, 54. doi: 10.1038/s41438-018-0053-y

Myouga, F., Akiyama, K., Motohashi, R., Kuromori, T., Ito, T., Iizumi, H., et al. (2010). The chloroplast function database: a large-scale collection of arabidopsis Ds/Spm- or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. *Plant J.* 61 (3), 529–542. doi: 10.1038/s41438-018-0053-y

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 (4-5), 447–460. doi: 10.1007/s11103-012-9873-6

Pogson, B. J., Ganguly, D., and Albrecht-Borth, V. (2015). Insights into chloroplast biogenesis and development. *Biochim. Biophys. Acta* 1847 (9), 1017–1024. doi: 10.1016/j.bbabio.2015.02.003

Qin, D., Dong, J., Xu, F., Guo, G., Ge, S., Xu, Q., et al. (2015). Characterization and fine mapping of a novel barley stage green-revertible albino gene (*HvSGRA*) by bulked segregant analysis based on SSR assay and specific length amplified fragment sequencing. *BMC Genomics* 16, 838. doi: 10.1186/s12864-015-2015-1

Qin, G., Gu, H., Ma, L., Peng, Y., Deng, X. W., Chen, Z., et al. (2007). Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res.* 17 (5), 471–482. doi: 10.1038/cr.2007.40

Qiu, Z., Kang, S., He, L., Zhao, J., Zhang, S., Hu, J., et al. (2018). The newly identified *heat-stress sensitive albino 1* gene affects chloroplast development in rice. *Plant Sci.* 267, 168–179. doi: 10.1016/j.plantsci.2017

Ramel, F., Ksas, B., Akkari, E., Mialoundama, A. S., Monnet, F., Krieger-Liszkay, A., et al. (2013). Light-induced acclimation of the arabidopsis chlorina1 mutant to singlet oxygen. *Plant Cell* 25 (4), 1445–1462. doi: 10.1105/tpc.113.109827

Roxas, V. P., Smith, R. K.Jr., Allen, E. R., and Allen, R. (1997). Overexpression of glutathione s-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. *Nat. Biotechnol.* 15 (10), 988–991. doi: 10.1038/nbt1097-988

Ruppel, N. J., Logsdon, C. A., Whippo, C. W., Inoue, K., and Hangarter, R. P. (2011). A mutation in arabidopsis seedling plastid development1 affects plastid differentiation in embryo-derived tissues during seedling growth. *Plant Physiol.* 155 (1), 342–353. doi: 10.1104/pp.110.161414

Sakowska, K., Alberti, G., Genesio, L., Peressotti, A., Delle Vedove, G., Gianelle, D., et al. (2018). Leaf and canopy photosynthesis of a chlorophyll deficient soybean mutant. *Plant Cell Environ.* 41 (6), 1427–1437. doi: 10.1111/pce.13180

Shanmugabalaji, V., Chahtane, H., Accossato, S., Rahire, M., Gouzerh, G., Lopez-Molina, L., et al. (2018). Chloroplast biogenesis controlled by DELLA-TOC159 interaction in early plant development. *Curr. Biol.* 28 (16), 2616–2623. doi: 10.1016/j.cub.2018.06.006

Shi, K., Gu, J., Guo, H., Zhao, L., Xie, Y., Xiong, H., et al. (2017). Transcriptome and proteomic analyses reveal multiple differences associated with chloroplast development in the spaceflight-induced wheat albino mutant *mta*. *PloS One* 12 (5), e0177992. doi: 10.1371/journal.pone.0177992

Su, N., Hu, M. L., Wu, D. X., Wu, F. Q., Fei, G. L., Lan, Y., et al. (2012). Disruption of a rice pentatricopeptide repeat protein causes a seedling-specific albino phenotype and its utilization to enhance seed purity in hybrid rice production. *Plant Physiol.* 159 (1), 227–238. doi: 10.1104/pp.112.195081

Sytykiewicz, H., Chrzanowski, G., Czerniewicz, P., Sprawka, I., Łukasik, I., Goławska, S., et al. (2014). Expression profiling of selected glutathione transferase genes in *Zea mays* (L.) seedlings infested with cereal aphids. *PloS One* 9 (11), e111863. doi: 10.1371/journal.pone.0111863

Trelease, R. N., Becker, W. M., Gruber, P. J., and Newcomb, E. H. (1971). Microbodies (Glyoxysomes and peroxisomes) in cucumber cotyledons: Correlative biochemical and ultrastructural study in light- and dark-grown seedlings. *Plant Physiol.* 48 (4), 461–475. doi: 10.1104/pp.48.4.461

Unlu, C., Drop, B., Croce, R., and van Amerongen, H. (2014). State transitions in chlamydomonas reinhardtii strongly modulate the functional size of photosystem II but not of photosystem I. *Proc. Natl. Acad. Sci. U.S.A.* 111 (9), 3460–3465. doi: 10.1073/pnas.1319164111

Wagner, R., Aigner, H., and Funk, C. (2012). FtsH proteases located in the plant chloroplast. *Plant Physiol.* 145 (1), 203–214. doi: 10.1111/j.1399-3054.2011.01548.x

Wang, Y., Wang, C., Zheng, M., Lyu, J., Xu, Y., Li, X., et al. (2016). *WHITE PANICLE1*, a Val-tRNA synthetase regulating chloroplast ribosome biogenesis in rice, is essential for early chloroplast development. *Plant Physiol.* 170 (4), 2110–2123. doi: 10.1104/pp.15.01949

Wang, C., Xu, W., Jin, H., Zhang, T., Lai, J., Zhou, X., et al. (2016). A putative chloroplast-localized Ca(2+)/H(+) antiporter CCHA1 is involved in calcium and pH homeostasis and required for PSII function in arabidopsis. *Mol. Plant* 9 (8), 1183–1196. doi: 10.1016/j.molp.2016.05.015

Wang, Y., Zhang, J., Shi, X., Peng, Y., Li, P., Lin, D., et al. (2016). Temperature-sensitive albino gene *TCD5*, encoding a monooxygenase, affects chloroplast development at low temperatures. *J. Exp. Bot.* 67 (17), 5187–5202. doi: 10.1093/jxb/erw287

Xing, S., Miao, J., Li, S., Qin, G., Tang, S., Li, H., et al. (2010). Disruption of the *1-deoxy-D-xylulose-5-phosphate reductoisomerase* (*DXR*) gene results in albino, dwarf and defects in trichome initiation and stomata closure in arabidopsis. *Cell Res.* 20 (6), 688–700. doi: 10.1038/cr.2010.54

Xiong, D., Huang, J., Peng, S., and Li, Y. (2017). A few enlarged chloroplasts are less efficient in photosynthesis than a large population of small chloroplasts in arabidopsis thaliana. *Sci. Rep.* 7 (1), 5782. doi: 10.1038/s41598-017-06460-0

Yang, J., and Cheng, Q. (2004). Origin and evolution of the *light-dependent protochlorophyllide oxidoreductase* (*LPOR*) genes. *Plant Biol. (Stuttg)* 6 (5), 537–544. doi: 10.1055/s-2004-821270

Ye, L. S., Zhang, Q., Pan, H., Huang, C., Yang, Z. N., and Yu, Q. B. (2017). *EMB2738*, which encodes a putative plastid-targeted GTP-binding protein, is essential for embryogenesis and chloroplast development in higher plants. *Physiol. Plant* 161 (3), 414–430. doi: 10.1111/ppl.12603

Yin, T., Pan, G., Liu, H., Wu, J., Li, Y., Zhao, Z., et al. (2012). The chloroplast ribosomal protein L21 gene is essential for plastid development and embryogenesis in arabidopsis. *Planta* 235 (5), 907–921. doi: 10.1007/s00425-011-1547-0

Zaltsman, A., Ori, N., and Adam, Z. (2005). Two types of FtsH protease subunits are required for chloroplast biogenesis and photosystem II repair in arabidopsis. *Plant Cell* 17 (10), 2782–2790. doi: 10.1105/tpc.105.035071

Zeng, X., Tang, R., Guo, H., Ke, S., Teng, B., Hung, Y. H., et al. (2017). A naturally occurring conditional albino mutant in rice caused by defects in the plastid-localized OsABCI8 transporter. *Plant Mol. Biol.* 94 (1-2), 137–148. doi: 10.1007/s11103-017-0598-4

Zhang, T., Feng, P., Li, Y., Yu, P., Yu, G., Sang, X., et al. (2018). *VIRESCENT-ALBINO LEAF 1* regulates leaf colour development and cell division in rice. *J. Exp. Bot.* 69 (20), 4791–4804. doi: 10.1093/jxb/ery250

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 (7780), 722–725. doi: 10.1038/s41586-019-1685-2

Zhao, D. S., Zhang, C. Q., Li, Q. F., Yang, Q. Q., Gu, M. H., and Liu, Q. Q. (2016). A residue substitution in the plastid ribosomal protein L12/AL1 produces defective plastid ribosome and causes early seedling lethality in rice. *Plant Mol. Biol.* 91 (1-2), 161–177. doi: 10.1007/s11103-016-0453-z

Zhu, X., Liang, S., Yin, J., Yuan, C., Wang, J., Li, W., et al. (2015). The DnaJ OsDjA7/8 is essential for chloroplast development in rice (*Oryza sativa*). *Gene* 574 (1), 11–19. doi: 10.1016/j.gene.2015.07.067

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 November 2022
DOI 10.3389/fpls.2022.1069618

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Changming Chen,
South China Agricultural University, China
Jianbin Hu,
Henan Agricultural University, China

*CORRESPONDENCE
Haibin Wu
wuhbhope@yeah.net
Gangjun Zhao
zhaogangjun@gdaas.cn

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

RECEIVED 14 October 2022
ACCEPTED 02 November 2022
PUBLISHED 18 November 2022

CITATION
Liu L, Gan Y, Luo J, Li J, Zheng X, Gong H, Liu X, Deng L, Zhao G and Wu H (2022) QTL mapping reveals candidate genes for main agronomic traits in *Luffa* based on a high-resolution genetic map. *Front. Plant Sci.* 13:1069618. doi: 10.3389/fpls.2022.1069618

QTL mapping reveals candidate genes for main agronomic traits in *Luffa* based on a high-resolution genetic map

Lili Liu[1], Yaqin Gan[1,2], Jianning Luo[1,3], Junxing Li[1], Xiaoming Zheng[1], Hao Gong[1], Xiaoxi Liu[1], Liting Deng[1], Gangjun Zhao[1]* and Haibin Wu[1,3]*

[1]Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [2]College of Coastal Agricultural Sciences, Guangdong Ocean University, Zhanjiang, China, [3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China

Luffa is an important medicinal and edible vegetable crop of Cucurbitaceae. Strong heterosis effects and strikingly complementary characteristics were found between the two domesticated *Luffa* cultivars, *Luffa acutangula* and *Luffa cylindrica*. To explore the genetic basis underlying their important agronomic traits, we constructed the first interspecific high-density genetic linkage map using a BC_1 population of 110 lines derived from a cross between S1174 (*Luffa acutangula*) and P93075 (*Luffa cylindrica*). The map spanned a total of 2246.74 cM with an average distance of 0.48 cM between adjacent markers. Thereafter, a large-scale field-based quantitative trait loci (QTLs) mapping was conducted for 25 important agronomic traits and 40 significant genetic loci distributed across 11 chromosomes were detected. Notably, a vital QTL (*qID2*) located on chromosome 9 with a minimum distance of 23 kb was identified to be responsible for the internode diameter and explained 11% of the phenotypic variation. *Lac09g006860* (*LacCRWN3*), encoding a nuclear lamina protein involved in the control of nuclear morphology, was the only gene harbored in *qID2*. Sequence alignment showed completely different promoter sequences between the two parental alleles of *LacCRWN3* except for some nonsynonymous single nucleotide polymorphisms (SNPs) in exons, and the expression level in thick-stem P93075 was distinctively higher than that in thin-stem S1174. According to the natural variation analysis of a population of 183 inbred lines, two main haplotypes were found for *LacCRWN3*: the P93075-like and S1174-like, with the former haplotype lines exhibiting significantly thicker internode diameters than those of the latter haplotype lines. It showed that *LacCRWN3*, as the only CRWN3 gene in Cucurbitaceae, was the most likely candidate gene regulating the internode diameter of *Luffa*. Our findings will be

beneficial for deciphering the molecular mechanism of key phenotypic traits and promoting maker-assisted breeding in *Luffa*.

KEYWORDS

***Luffa*, genetic linkage map, vital agronomic traits, QTL mapping, candidate genes, haplotype analysis**

Introduction

Luffa is a monoecious and cross-pollinated diploid belonging to the Cucurbitaceae family (Wu et al., 2016). The *Luffa* genus consists of nine species (Filipowicz and Schaefer, 2014), among which *Luffa acutangula* (Ridge gourd) and *Luffa cylindrica* (Sponge gourd) are domesticated and cultivated worldwide, especially in India, China, Thailand, Central America and Africa (Rabei et al., 2013; Wu et al., 2014). Generally, young *Luffa* fruit commercially produced as an edible vegetable is rich in nutrients, such as proteins, citrulline, vitamins, trace elements and crude fiber (Tyagi et al., 2020). The fruit is also rich in bioactive compounds for medicinal treatment, such as flavonoids, alkaloids, sterols and glycosides (Gowtham et al., 2012; Abdel-Salam et al., 2019). For the mature/old fruit, it includes a tightly arranged network composed of large amounts of fiber, which can be utilized as cleaning products, packaging and industrial materials (Zhang et al., 2007; Papanicolaou et al., 2015). Therefore, *Luffa* should be well-studied as a commercially cultivated crop with nutritional, medicinal, and industrial values.

Several strikingly complementary characteristics have been identified between the two domesticated cultivars: *Luffa cylindrica* has smooth and cylindrical fruit, strong adaptability and disease resistance; while *Luffa acutangular* has ridge gourd fruit, small pale-yellow flowers, good fruit quality and performances during storage and transportation. Furthermore, it is reported that strong heterosis effects exist between the two domesticated cultivars (Wu et al., 2016). However, under natural conditions, the sponge gourd's flowers usually open between 4:00 and 6:00 a.m., while the ridge gourd bloom between 4:00 and 6:00 p.m. The nearly 12-h gap in the daily flowering time hinders their natural hybridization and gene flow. Thence, in-depth investigation of the genetic and molecular basis of these phenotypic traits in both cultivars is of great significance to promote the utilization of interspecific heterosis and the breeding of new *Luffa* varieties with excellent agronomic performances.

The cucurbit model plant cucumber has been extensively studied to dissect the genetic architecture of important agronomic morphology (Pan et al., 2020a; Pan et al., 2020b; Chen et al., 2021; Shan et al., 2021). In addition to a large number of identified quantitative trait loci (QTLs) (Pan et al., 2020a; Wang et al., 2020), certain genes for corresponding traits, including fruit shape (Jiang et al., 2015; Xin et al., 2019; Zhao et al., 2019; Zhang et al., 2020b), inflorescence architecture (Wen et al., 2021), flowering time (Cai et al., 2020), plant height (Wang et al., 2017a) and stem diameter (Yang et al., 2022), have been characterized, which have considerably facilitated the genetic and molecular breeding of cucumbers. Additionally, a series of QTLs responsible for key phenotypic traits have also been detected in other cucurbits, including melon and watermelon (Ren et al., 2018; Guo et al., 2020; Pan et al., 2020b; Cao et al., 2021).

In comparison, scientists have mainly focused on the traditional *Luffa* breeding program in past decades, and the genetic and molecular research is relatively insufficient. Wu et al. (2016) constructed an interspecific genetic linkage map between *Luffa acutangula* and *Luffa cylindrica* based on 177 expressed sequence tag-derived simple sequence repeat (EST-SSR) markers distributed in 14 linkage groups (LGs), which spanned 1436.12 cM with an average of 102.58 cM per LG, and six putative QTLs associated with reproductive isolation traits were identified. The recent publications on the genomes of *Luffa acutangul*a and *Luffa cylindrica* (Wu et al., 2020; Zhang et al., 2020a; Pootakham et al., 2021) provide useful resources for developing high-throughput genomic variations to improve the efficiency of loci or gene mining. Lou et al. (2020) constructed a high-density genetic map with the genetic background of *Luffa cylindrica*, in which 3701 polymorphic markers based on specific-locus amplified fragment sequencing (SLAF-seq) were mapped to 13 LGs and the map spanned a total distance of 1518.56 cM. Using this genetic map, a LTR Copia-type retrotransposon harbored in the target QTL was predicted as the candidate gene responsible for CMV resistance. Through bulk-segregant analysis sequencing (BSA-Seq) and comparative genomics analyses, a dwarfism gene encoding a GA3ox was mapped using an F_2 *Luffa acutangular* population, and the large insertion in the mutant allele might account for the defective GAs biosynthesis and dwarf phenotype in WJ209 (Zhao et al., 2021). Moreover, transcriptome and sRNAome, and degradome have also been used for revealing network associated with fruit skin coloration in *Luffa cylindrica* (Sun et al., 2022). To our knowledge, based on the interspecific genetic background between *Luffa acutangula* and *Luffa cylindrica*, no studies have been reported on the comprehensive and high-resolution analyses of the genetic architecture for key agronomic traits.

In the current study, a high-density genetic linkage map across *Luffa* interspecies was constructed using SLAF-seq-based single nucleotide polymorphism (SNP) markers. QTL mapping for important agronomic traits was further performed and key candidate genes were evaluated. In particular, a key candidate gene (*LacCRWN3*) controlling the internode diameter was characterized, and its differential expression profiles associated with the sequence variations in promoter regions might be the main causation for stem thickness divergence between the thin-stem parental line S1174 and thick-stem parental line P93075. Meanwhile, two distinct haplotypes of *LacCRWN3* were found to be responsible for the internode diameter variation in natural population. This study will facilitate the molecular research and maker-assisted breeding in *Luffa*.

Materials and methods

Plant materials and growing conditions

The BC_1 mapping population comprising 110 individuals was derived from a cross between inbred lines P93075 (*Luffa cylindrica*) and S1174 (*Luffa acutangula*), with female line S1174 as the recurrent parent. *Luffa cylindrica* and *Luffa acutangular* are two domesticated *Luffa* cultivars that could be distinguished by different characteristics. The former has smooth and cylindrical fruit, and bright yellow flowers that open early in the morning, whereas the latter has ridged gourd fruit with flowers that bloom late in the afternoon (Supplementary Figure S1).

Additionally, a core panel of 183 *Luffa* inbred lines was used for haplotype analysis of the candidate gene. The panel included 94 *Luffa acutangular* and 89 *Luffa cylindrica*, exhibiting abundant genetic and phenotypic diversity.

Field experiments were carried out in Baiyun Field Trial Base of Guangdong Academy of Agricultural Sciences (23° 15′ N, 113° 27′ E) during the spring-summer season of the year 2015 for the BC_1 population, and the field management followed local practices. The parental and offspring lines were grown in a completely randomized design across eight rows of 20 individuals. Each row was 8 m long with a row spacing of 0.4 m. For the 183 inbred lines, the field trials were conducted in the same location in the summer-autumn of 2021 and spring-summer of 2022. Each inbred line was grown in two rows, with 3 m long and 0.7 m apart, and 12 individuals were planted for each line.

Phenotyping and phenotypic data analysis

During the growing season of *Luffa*, 110 BC_1 individuals and their parental lines were investigated for four categories of agronomic variables: maturity-related traits, flower-size-related traits, plant-growth-related traits and fruit-related traits. At the flowering stage, the first node for the male flower (FMFN), female flower (FFFN) and fruit (FFRN), daily flowering time (FT) and male flower diameter (MFD) were measured to evaluate plant maturity and flower size. During the late stage of plant growth, the main stem and branch-related traits were determined, including the total number of internodes (TNI), length and diameter of the 7th, 8th and 9th internodes (IL_7th, IL_8th, IL_9th, IL_AV, ID_7th, ID_8th, ID_9th, ID_AV), growth rate (GR), branch number (BN) and total branch length (TBL). During the fruit development stage, age-related parameters such as fruit length and diameter (YFL, OFL, YFD, OFD), as well as fruit peduncle length and diameter (YFPL, OFPL, YFPD, OFPD), were measured. For the 183 inbred lines, only the internode diameter was examined. A description of the phenotypic traits is presented in Supplementary Table S1.

All phenotypic statistical analyses were performed in R (v3.4.2). The Pearson correlation coefficients among different traits were computed and visualized by the Corrplot and Hmisc packages. The Skewness and Kurtosis functions in the fBasics package were used for the normal test. Phenotypic frequency distributions were plotted using OriginPro 9.0.0.

Genomic DNA extraction

The genomic DNA was extracted from three-week-old leaves of all samples, including the two parental lines and their 110 BC_1 progeny individuals, as well as the 183 inbred lines. DNA extraction was performed using the modified cetyltrimethylammonium bromide (CTAB) method, and then purification with a plant genomic DNA extract kit (Tiangen Inc., Beijing, China). DNA concentration and quality were measured by NanoDrop ND-1000 Spectrophotometer and 1% (w/v) agarose gel electrophoresis.

SLAF-seq and genotyping

The strategy of SLAF-seq described by Sun et al. (2013) was utilized for the BC_1 population and its parental lines with some modifications. Briefly, the cucumber(http://www.icugi.org/cgi-bin/ICuGI/genome/home.cgi?organism=cucumber&ver=2)

was selected as the reference genome for predicting enzyme digestion due to a lack of the *Luffa* genome in 2016, and the restriction enzymes of HaeIII+Hpy166II were specified to digest the genomic DNA. Afterwards, the fragments ranging from 314 to 364 bp were purified to construct the SLAF library and perform pair-end sequencing (Each end 125 bp) on an Illumina HiSeq 2500 platform (Illumina, Inc., San Diego, CA, USA) by Biomarker Technologies Corporation (Beijing, China). A control experiment with *Oryza sativa ssp. japonica* was

conducted to evaluate the accuracy and reliability of the SLAF library construction and sequencing. The results showed that the enzyme digestion efficiency of control data was 91.76%, and the paired-end mapping efficiency was 78.56%, indicating that the experimental process was effective. After trimming adapters and filtering low quality raw reads of SLAF (Reads with > 10% unidentified nucleotides [N]; > 50% bases having Phred quality score ≤ 5), the clean reads were mapped to reference genome *Luffa acutangular* S1174 (Unpublished data) using Burrows Wheeler Alignment (BWA, v0.7.8) (Li and Durbin, 2009). For the uniquely mapped reads, SNP calling was performed with Genome Analysis Toolkit (GATK, v3.2-2) (Mckenna et al., 2010), and the detected SNPs were designated as the final variations for subsequent analysis.

To develop available molecular markers for construction genetic linkage map, we first identified SNPs that were bi-allelic, homozygous and polymorphic between parental lines, and a set of targeted SNPs of aa × bb type was identified. Furthermore, the genotypes of 110 BC_1 individuals at these polymorphic loci were extracted and filtered according to the following criteria: SNPs with multiple alleles that were not inherited from parents were removed; SNPs genotyped in more than 99% of offspring individuals were retained; SNPs with significant segregation distortion (χ^2 test, $P < 0.001$, D.F. = 2) were excluded. Finally, a collection of high-quality SNP markers was obtained for linkage map construction.

Genetic linkage map construction and QTL mapping

The *Luffa* genetic map was constructed by the Lep-MAP3 software, with the maximum likelihood method employed to order the markers within each LG, and Kosambi mapping function used for converting recombination percentages to genetic distances in cM (Rastas et al., 2017).

QTL analysis was performed using the composite interval mapping (CIM) model by WinQTL Cartographer software (v2.5) (Wang et al., 2012). The logarithm of odds (LOD) threshold for QTL detection was determined based on 1000 permutations at a significance level of $P < 0.05$ for every trait with the function of permutation test in MapQTL (v6.0) (Van Ooijen, 2009). At the highest probability peak, the phenotypic variance explained by each QTL was achieved.

Candidate genes identification and gene ontology analysis

All the genes within the QTLs were identified as potential candidate genes. The most likely genes were further selected according to their functional annotation and orthologous gene prediction in other species. For the candidate genes of a certain trait, gene ontology (GO) enrichment analysis was performed using Singular Enrichment Analysis tool in agriGO1.2 (Tian et al., 2017). A false discovery rate of 0.05 was used to identify significant GO terms.

Haplotype analysis of the candidate gene

Whole genome re-sequencing of the 183 core *Luffa* lines was performed to examine the haplotype distribution of the candidate gene. The following is a summary of the re-sequencing and SNP detection pipeline: sequencing was done on the DNBSEQ platform (China National GeneBank, Shenzhen, China) by 150 bp paired-end sequencing with an average depth of approximately 6.48 ×. Adapters were trimmed and low quality reads (Reads with > 10% unidentified nucleotides [N]; > 50% bases having Phred quality score ≤ 5) were filtered before mapping the clean reads to the S1174 reference genome (Unpublished data) using BWA (v0.7.12) (Li and Durbin, 2009). SNP and insertion and deletion (InDel) calling were performed with GATK (v4.2.2.0) (Mckenna et al., 2010), and ANNOVAR was used for functional annotation (Wang et al., 2010). The obtained variations were further filtered with a missing rate < 0.5 and a minor allele frequency > 0.05. Subsequently, we extracted the SNPs and InDels around the candidate gene in each *Luffa* accession to examine the haplotypes.

Phylogenetic tree construction

Through sequence alignment, two CRWN homologs (CRWN3 and CRWN4) were found in *Luffa*, and the amino acid sequences similar to both proteins were downloaded from public databases of maize (*Zea mays*), rice (*Oryza sativa*), *Arabidopsis thaliana* and other eight Cucurbitaceae crops. The sequences were aligned by ClustalW, and a maximum-likelihood phylogenetic tree was constructed with 1000 bootstrap replicates using MEGA (v7.0.26). Online software MEME was used to predict the protein motifs with a maximum of 10 motifs and a maximum width of 100.

RNA extraction and quantitative real-time PCR

Total RNA was extracted from the 7th, 8th and 9th internode of main stem in three biological replications for each parent, with a quick RNA isolation kit (TransGen Biotech, China), and complementary DNA (cDNA) was synthesized by the Transcript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, China) using a 20 µl reaction system. The quantitative real-time PCR (qRT-

PCR) was performed using TB Green Premix Ex Taq II (Takara, Japan) in three replications for each sample. *Luffa 18S rRNA* gene was used as an internal control and relative expression was calculate by $2^{-\Delta\Delta ct}$ method (Livak and Schmittgen, 2001).

Primers used for qRT-PCR were listed as below: the *18S rRNA* gene (F: 5'-GTGTTCTTCGGAATGACTGG-3', R: 5'-ATCGTTTACGGCATGGACTA-3'); the candidate gene *Lac09g006860* in S1174 (F: 5'-CGTAATGGAG CGCAAAGATCGG-3', R: 5'-CTTGCCCTAGTTGGTCATACTT-3') and P93075 (F: 5'- CGTAATGGAGCGCAAAGATCGG -3', R: 5'-TCTTGCCCTAGTTGGTCGTACTT -3').

Statistical analysis

One-way ANOVA followed by Duncan's or Dunnett's T3 multiple range tests ($P < 0.05$) or two-tailed student's *t* test ($P < 0.05$) were performed to evaluate the data differences by SPSS 18.0.

Results

Analysis of high-throughput sequencing data and SNP markers development

After SLAF-seq and raw reads filtering, a total of 2.33 G, 2.11 G and 39.75 G clean data were obtained for the maternal and paternal parents and BC_1 population. Each offspring individual's average data was greater than 0.36 G, with the average Q30 of 91.59% and guanine cytosine (GC) content of 40.68% (Supplementary Table S2). The average mapping rate of all progeny samples was 96.46%, and their average sequencing depth was 1.45 ×, ranging from 1.24 × to 1.71 × (Supplementary Table S2). The results indicated that the sequencing quality was high and the data could be used for subsequent analysis.

Through statistical analysis of SNP data between paternal and maternal parents, 388,402 polymorphic SNPs following the genetics' two allelic general encoding rules were developed altogether, and 319,601 fell into the aa × bb segregation pattern (Supplementary Table S3). The genotypes of 110 BC_1 individuals were identified at these targeted polymorphic loci and further filtered with criteria described in method section. Ultimately, a collection of 4,798 polymorphic SNP markers was obtained and used for linkage map construction.

Construction of high-resolution linkage map for *Luffa*

By using the Lep-MAP3 software, 4,673 markers were successfully mapped onto the linkage map of BC_1 population, resulting in an effective mapping ratio of available SNPs approximately 97.58%. The genetic map included 13 LGs consistent with the haploid chromosome number of *Luffa*, spanning a total of 2246.74 cM with an average distance of 0.48 cM between adjacent markers (Figure 1; Supplementary Table S4). The genetic length within LGs varied from 144.09 cM in LG11 to 235.25 cM in LG12, with an average of 172.83 cM, and the number of markers ranged from 237 in LG11 to 481 in LG1, with an average of 359.46 (Supplementary Table S4). LG1 was the densest group, with a total distance and an average marker interval of 160.45 cM and 0.33 cM, respectively (Supplementary Table S4). The number of markers with a gap less than 5 cM accounted for 97.15% of the linkage map, indicating relatively good distribution uniformity. Since the genome size of *Luffa acutangular* S1174 has been estimated to be 776.49 Mb (Unpublished data), the average recombination rate across all LGs was approximately 2.89 cM/Mb. This is the highest density genetic map of *Luffa* reported to date (Figure 1), which is generally considered suitable for QTL mapping.

Phenotypic variation analysis

Significant variations were observed for all phenotypes in the two parental lines (Table 1). P93075 showed higher first node for male flower (FMFN), female flower (FFFN) and fruit (FFRN), and earlier diurnal flowering time (FT) than those of S1174, (Table 1; Supplementary Figures S2, S3), indicating that S1174 was an early-maturing variety in terms of overall plant growth and development. For the plant-growth related traits, S1174 exhibited lower growth rate (GR), total number of internodes (TNI), branch number (BN), total branch length (TBL), and internode diameter (ID) (Table 1; Supplementary Figures S2, S3). The results suggested that the whole plant growth momentum of P93075 was more vigorous than that of S1174, whereas the latter might perform bettered in the 7th, 8th and 9th internode lengths (ILs). The fruit length of S1174 was longer compared to that of P93075 during the whole fruit development period, and its fruit diameter was larger at the young fruit stage but thinner at the old fruit stage (Table 1; Supplementary Figures S2, S3). Meanwhile, distinct phenotypic variations in the derived BC_1 population were also investigated, with the coefficient of variation (*CV*) ranging from 5.52 in FT to 82.14 in TBL (Table 1).

The skewness and kurtosis test demonstrated that all the traits were approximately normally distributed, highlighting that they followed the genetic characteristics of a quantitative trait and might be controlled by multiple genes (Table 1; Supplementary Figures S2, S3). Pearson correlation analysis showed that there were significant correlations among many traits within one category, and the closest correlations were found among IL and ID (0.83-0.90, $P < 0.001$), followed by that between FFFN and FFRN (0.83, $P < 0.001$) (Supplementary

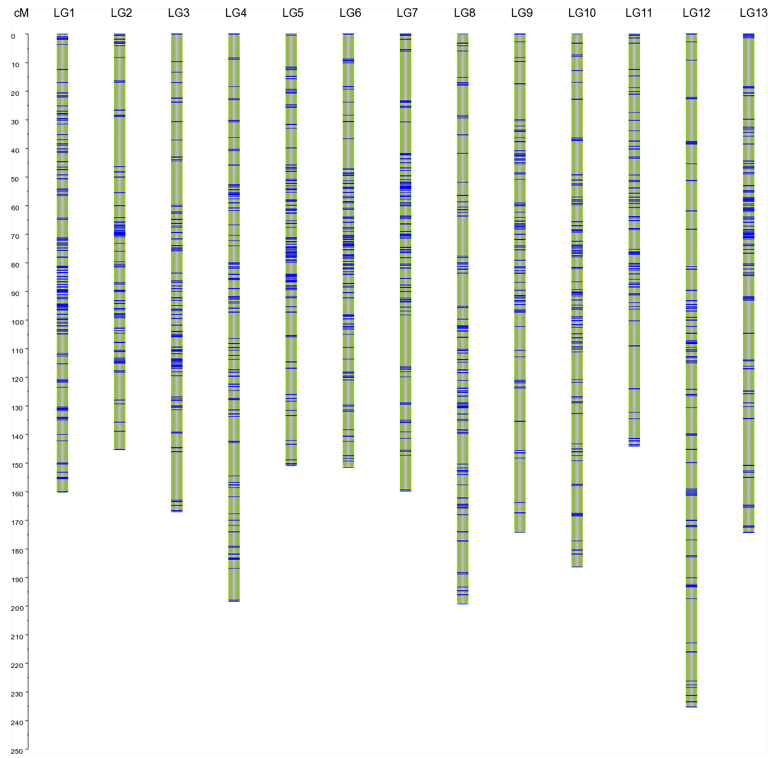


FIGURE 1
An interspecific high-density genetic linkage map of *Luffa* with 13 linkage groups. Single nucleotide polymorphism (SNP) markers are highlighted with blue lines on each linkage group (LG).

Figures S4A, C). As expected, both length and diameter between fruit and fruit peduncle were positively correlated, and TNI was also found to be positively correlated with BN and TBL (Supplementary Figures S4B, C). While for traits across different categories, the strongest correlations were found between FFFN, FFRN and TNI with the r of 0.34 ($P < 0.01$) (Supplementary Table S5).

QTL mapping

Totally, 40 QTLs associated with 17 traits were distributed across 11 chromosomes of *Luffa* except chromosomes 2 and 3, and the LOD scores ranged from 3.46 to 23.12, while no QTL was detected for eight traits including TNI, BN, IL_AV, ID_7th, ID_8th, ID_9th, old fruit diameter (OFD) and old fruit peduncle diameter (OFPD) (Figure 2; Supplementary Table S6). The estimated phenotypic variation explained (PVE) by the significant locus individually varied from 5% to 47%, with 80% of them accounting for > 10% and 15 loci > 15%, suggesting that they were the major effect QTLs responsible for the genetic controls of these traits. Among the 40 QTLs, the additive effects of 10 QTLs were derived from S1174 alleles, whereas those of the other 30 QTLs were from P93075 alleles.

For maturity-related traits, 12 QTLs were mapped to chromosomes 1, 7, 9, 10, 11 and 13 with PVE values of 5-26% (Figure 2; Supplementary Table S6). QTL *qFT5* had the largest effect, which explained a phenotypic variation of 26%, and the allele from P93075 was correlated with 1.17 h advanced in daily flowering time. The followed largest effect QTLs, *qFFRN1* and *qFFFN1* with PVEs of 18% and 17.2%, were located on chromosomes 7 and 1, and the allele from S1174 at the corresponding locus was responsible for a decrease in FFRN and FFFN of 11.41 and 10.58, respectively. *qFFRN2*, another large-effect QTL with ≥ 15% PVE, located on chromosome 11, and the S1174 allele contributed to an increase of 9.8 in FFRN.

Meanwhile, three QTLs on chromosomes 5 and 7 were identified for male flower dimeter, accounting for 40% of the total phenotypic variance, and those alleles with increasing effect were all from P93075 (Figure 2; Supplementary Table S6).

For plant-growth-related traits, nine QTLs were found localized on chromosomes 5, 7, 8, 9, 10 and 11, with PVEs

TABLE 1 Basic description of phenotypic data in parental lines and the BC1 population.

Category	Traits	Parents		BC1 lines					
		S1174	93075	Min	Max	Mean ± SE	Skewness	Kurtosis	*CV*
Maturity related traits	FMFN	2.50	16.56	1.00	12.00	4.10 ± 0.20	1.45	1.98	51.07
	FFFN	6.00	33.00	4.00	65.00	20.74 ± 1.18	1.26	1.97	59.51
	FFRN	9.23	34.50	7.00	66.00	26.85 ± 1.21	0.90	0.63	47.15
	FT	17.17	5.00	17.87	22.62	19.72 ± 0.10	1.01	0.83	5.52
Flower size related traits	MFD (mm)	8.63	14.79	5.60	12.01	9.00 ± 0.12	0.18	-0.03	14.01
Plant growth related traits	GR	0.89	0.90	0.50	2.17	1.30 ± 0.03	0.07	0.14	24.94
	TNI	27.33	29.50	21.00	44.00	32.28 ± 0.42	-0.29	-0.09	13.51
	IL_7th (cm)	12.92	8.91	2.50	19.30	10.97 ± 0.33	-0.19	-0.16	31.43
	IL_8th (cm)	12.46	8.94	3.00	20.30	11.40 ± 0.36	0.14	-0.38	33.16
	IL_9th (cm)	13.02	9.77	3.00	20.80	11.78 ± 0.33	0.15	-0.06	29.23
	IL_AV (cm)	12.80	9.21	5.07	18.17	11.38 ± 0.29	0.10	-0.58	26.72
	ID_7th (mm)	4.55	4.75	3.00	10.50	5.57 ± 0.09	1.12	5.33	17.34
	ID_8th (mm)	4.77	4.82	3.00	10.30	5.47 ± 0.09	1.11	5.14	17.55
	ID_9th (mm)	4.67	4.91	3.40	9.80	5.37 ± 0.08	1.08	6.22	15.47
	ID_AV (mm)	4.66	4.83	3.13	8.07	5.47 ± 0.07	0.08	0.91	14.29
	BN	1.85	5.10	1.00	13.00	4.55 ± 0.28	0.80	0.06	63.53
	TBL (cm)	103.30	171.05	0.00	568.00	159.67 ± 12.51	0.92	0.34	82.14
Fruit related traits	YFL (cm)	41.00	31.20	11.80	64.50	42.65 ± 0.86	-0.24	0.58	21.04
	YFD (mm)	64.02	63.80	33.00	76.70	60.07 ± 0.73	-0.35	0.88	12.80
	YFPL (cm)	12.00	14.07	6.80	47.40	12.98 ± 0.42	4.63	35.56	33.59
	YFPD (mm)	9.20	10.34	7.40	15.40	11.13 ± 0.15	0.13	-0.22	14.39
	OFL (cm)	47.38	40.60	29.80	82.70	48.59 ± 0.87	0.71	1.23	18.86
	OFD (mm)	63.28	98.28	34.80	110.00	75.70 ± 1.22	-0.58	1.14	16.96
	OFPL (cm)	11.13	13.70	6.50	21.80	12.59 ± 0.27	0.85	1.34	22.49
	OFPD (mm)	11.95	11.44	3.20	16.10	11.40 ± 0.22	-0.91	1.16	19.94

ranging from 10% to 23%, of which the increasing effects for *qGR1*, *qIL3*, *qIL4*, *qID2* and *qTBL1* were from P93075, and those for other QTLs of *qID1*, *qIL1*, *qIL2* and *qIL5* from S1174 (Figure 2; Supplementary Table S6).

For fruit-related traits, 16 QTLs were detected across eight chromosomes, contributing to PVE values of 8%-47% (Figure 2; Supplementary Table S6). Among them, *qYFPL5* and *qYFPL2*, which are located on chromosomes 7 and 5, had the greatest effects and accounted for 47% and 40% of variance in young fruit peduncle length (OFPL). The P93075 allele at *qYFPL5* and S1174 allele at *qYFPL2* had the increased effects of YFPL. Regarding to old fruit peduncle length, two QTLs, *qOFPL2* and *qOFPL3* on chromosome 8, were detected, explaining 24% and 19% of total phenotypic variation, and the P93075 alleles at these two loci increased the OFPL by 2.79 cm and 2.52 cm, respectively. The three QTLs related to young fruit, *qYFD1* on chromosome 1, *qYFD2* on chromosome 10 and *qYFL1* on chromosome 5, accounted for ≥ 15% PVE, and the alleles from S1174, P93075 and P93075 were responsible for the increase in the corresponding traits of 6.29 mm, 7.32 mm and 7.34 cm, respectively.

Candidate genes analysis for the three key types of agronomic phenotypes

The genes underlying the six QTLs for daily flowering time, five QTLs for fruit shape and seven QTLs for internode development were retrieved, and a total of 2,980 genes were detected.

The six QTLs detected for daily flowering time were found to harbor 633 genes. Based on GO annotation and orthologous gene prediction, four key genes were prioritized to be the most likely genes (Supplementary Table S7). *Lac01g003380* and *Lac01g013890* were two genes enrichened in the circadian rhythm network for modulating flowering time *via* the photoperiod pathway, and they encoded a phytochrome C and a casein kinase II subunit beta (CK2β), respectively. Phytochrome C was reported as the photoreceptor to transmit signals to the circadian clock (Kircher et al., 2002), and CK2 acted as a ubiquitous Ser/Thr kinase to phosphorylate the central clock components CCA1 and LHY (Mulekar and Huq, 2012). The two *Arabidopsis* orthologs, *Lac01g012970* (*AtSEP3*) and *Lac01g012980* (*AtSOC1*), both encoded MADS-box

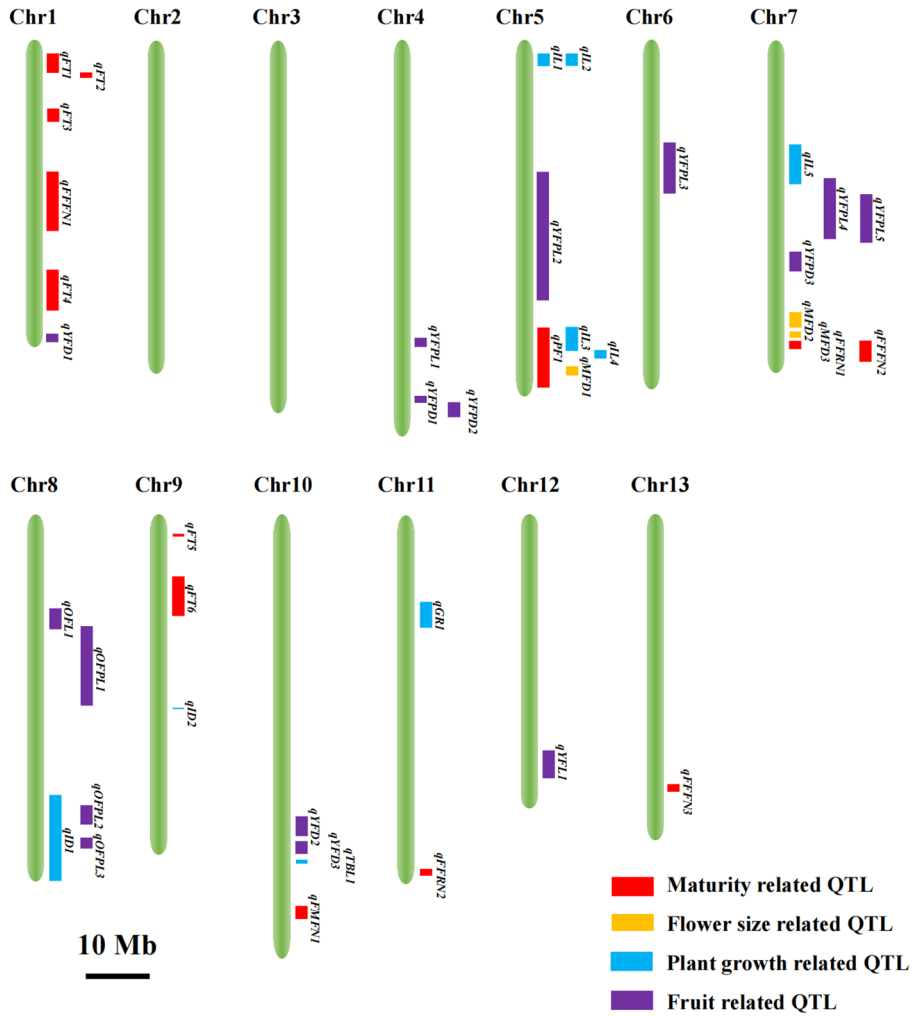


FIGURE 2
Visualization of the chromosomal locations of the 40 QTLs. The red, yellow, blue and purple rectangles represent the QTLs related to maturity, flower size, plant growth and fruit, respectively. The 10 Mb bar indicates the size of regions in the reference genome *Luffa acutangular* S1174. Chr, chromosome.

transcription factors, were also involved in the regulation of flowering time and flower morphogenesis.

A total of 731 genes were found underlying the five QTLs regions detected for fruit shape, and putative six genes were prioritized, three of which were enriched in cell cycle, two in plant hormone responses, and one selected for its significance in other crops (Supplementary Table S7). Two cyclins (*Lac10g017950* and *Lac10g018520*) and one CDK inhibitor gene (*Lac10g016980*) were shown to modulate the cell division and expansion of fruit by changing their expression patterns in cell cycle progression (Anwar et al., 2019). *Lac10g015340* and *Lac10g015350*, both encoding auxin-responsive protein IAA14, were predicted to play roles in cell enlargement during fruit development. *Lac10g015060* (*SUN*, *Cla011257*) encoding an IQ-DOMAIN14 protein was another most likely candidate gene regulating fruit shape.

For the five QTLs of internode length, 551 genes were detected and 11 were prioritized according to previous findings that the improved plant height of high-yielding varieties generated during the 'green revolution' was primarily due to the significant change of GA pathway, and its crosstalk with other phytohormones (Wang et al., 2017b). These candidate genes included *GA20ox* and *gibberellin receptor GID1B-like* from GA biosynthesis and signaling; *TIR1*, *auxin-induced protein AUX28-like* and *22D-like* from auxin metabolism; five ethylene-responsive transcription factors in ethylene signaling; and *TIFY 9-like isoform X2* from JA pathway (Supplementary Table S7).

In addition, Two QTLs for the internode diameter, *qID1* and *qID2*, were detected. We mainly focused on *qID2*, which was located on chromosome 9 with a minimum distance of 23 kb and explained 11% of the phenotypic variation (Figure 2; Supplementary Table S6). Fascinatingly, *qID2* harbored only one candidate gene, *Lac09g006860*, which encodes a CROWDED NUCLEI 3 (LacCRWN3) protein involved in the regulation of nuclear structure.

LacCRWN3 responsible for the internode diameter of main stem in *Luffa*

Lac09g006860 was an extremely large gene of 31,127 bp and its paralogue in P93075 was *Lcy09g015130* with a length of 53,482 bp, and their amino acid identity reached 98.07%. To detect genomic variations of this candidate gene between parental lines S1174 and P93075, DNA sequences containing the whole genomic regions and the 2000 bp upstream regions of ATG were extracted from the two cultivars' *de novo* data (Wu et al., 2020; Unpublished data). In the genomic region, 40 SNPs were present in exons, which caused 23 nonsynonymous mutations (Dittmer et al., 2007; Zhao et al., 2016). We further compared the upstream regulatory regions, and a total of six variations including four SNPs and two InDels were detected in 5' un-translated regions (UTRs), four of which were related to major cis-elements (E-box: 5′-CANNTG-3′; ARE motif) (Figure 3A). Compared with P93075, there were three base mutations around the MYC binding motif E-box (265-273 bp upstream of ATG) in S1174, and a 7 bp deletion (-86 bp) was also detected next to the ARE cis-acting regulatory element essential for the anaerobic induction. Interestingly, two alleles of the candidate gene had completely different promoter sequences upstream of 5' UTR, and many vital regulatory elements related to hormones and abiotic stress were identified. In particular, the unique ABRE motifs related to abscisic acid (ABA) responses were identified in P93075.

Considering that the sequences of this candidate gene *LacCRWN3* were distinctively different in the upstream regulatory regions, we hypothesized that the observed internode diameter variation might be primarily due to the differential gene expression patterns in parental lines. To

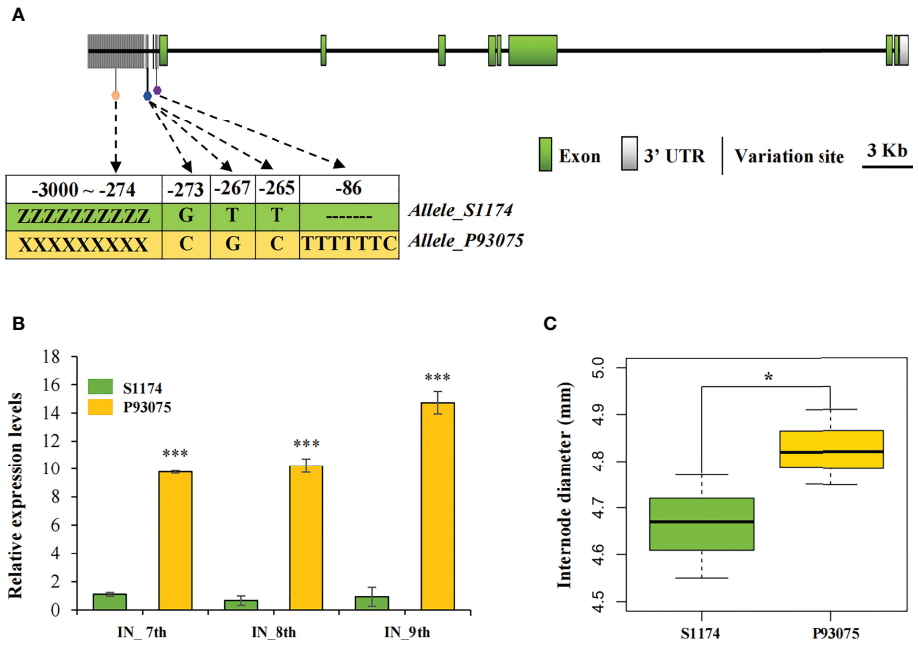


FIGURE 3
Candidate gene *LacCRWN3* targeted by *qID2* responsible for internode diameter. **(A)** Schematic representation of the structure and sequence variations of *LacCRWN3*. The white box represents the 3' un-translated region (UTR), and the green boxes represent exons. The upstream region before the first exon indicates the 5' UTR and promoter region. The vertical black lines indicate variations between S1174 and P93075 in the upstream region. The pentagons represent variation sites in the major cis-elements, different colors indicate different cis-elements. ZZZZZZZZZ and XXXXXXXXX refer to completely different allelic sequences in the region -3000 to -274 upstream of the start codon ATG. Variations of the marked sites between alleles from S1174 and P93075 are displayed in green and yellow backgrounds, respectively. **(B)** Expression profiles of *LacCRWN3* in the 7th, 8th, and 9th internode of parental lines S1174 and P93075. **(C)** Phenotypic variation of internode diameter between parental lines S1174 and P93075. Results indicate mean ± SE of three replicates, ***$P < 0.001$; *$P < 0.05$.

confirm this speculation, qRT-PCR assays were initially conducted in the 7th, 8th and 9th internode between S1174 and P93075, and the data showed that expression levels of *LacCRWN3* in P93075 were significantly higher than those in S1174 for all the three internodes (Figure 3B). Additionally, in P93075, *LacCRWN3* was strongly upregulated from the 7th to 9th internode, while in S1174, the expression levels were weak and maintained almost at the same level for all the three internodes (Figure 3B). The results indicated that differential gene expression profiles were tightly linked to the phenotypic variation of the two parental lines, and *LacCRWN3* may act as a positive regulator of internode diameter (Figures 3B, C).

Two haplotypes of *LacCRWN3* identified among the inbred line population

To further explore the functions of vital alleles of *LacCRWN3*, we genotyped natural allelic variations in 183 *Luffa* inbred lines by re-sequencing. Intriguingly, according to the SNPs and InDels covering the entire gene and 2000 bp upstream region of ATG, two main haplotypes of *LacCRWN3* were identified in the population, namely the S1174-like and P93075-like (Figure 4A). In total, 50 SNPs were detected in exon regions, among which 24 were nonsynonymous, and eight variations (two SNPs and six InDels) and completely different sequences were found in the 5'UTR and promoter regions, respectively. Statistically, lines with the P93075-like haplotype showed significantly ($P < 0.5$) thicker main stem diameters than those with the S1174-like haplotype (Figure 4B), further suggesting that the *P93075-like* allele can thicken the stem of *Luffa*. The results demonstrated that *LacCRWN3* was indeed a key candidate regulatory gene responsible for the variation of internode diameter in *Luffa*.

LacCRWN3 being the only *CRWN3* gene in Cucurbitaceae crops

It has been reported that the CRWN family consists of four members with each contains a large central coiled-coil domain (Dittmer et al., 2007; Wang et al., 2019). To investigate the potential diversification of CRWN family in Cucurbitaceae, the phylogenetic analysis was done (Figure 5). The results illustrated that there were three members of the CRWN in *Luffa*, including the CRWN1-like, CRWN3-like (LacCRWN3) and CRWN4-like, while other Cucurbitaceae species possessed only the CRWN1-like and CRWN4-like proteins. Therefore, our targeted LacCRWN3 was the only CRWN3 protein in Cucurbitaceae, which hinted that a special pathway mediated by CRWN3 to regulate the stem development may exist in *Luff* compared with the mechanisms in other Cucurbitaceae.

Discussion

Luffa is an important cucurbit vegetable crop grown worldwide. Strong heterosis effects and strikingly complementary characteristics were found between its two domesticated cultivars, *Luffa acutangular* and *Luffa cylindrica*. To explore the genetic basis underlying important agronomic traits of the two cultivars, we generated an interspecies BC_1 population of 110 individuals to construct the high-quality genetic map. Simultaneously, QTL mapping and candidate gene identification of important agronomic traits were carried out.

The most comprehensive dataset of phenotypic analysis in *Luffa*

The two parents of the BC_1 population, S1174 and P93075, are representative lines of the two commonly cultivated *Luffa* species, *Luffa acutangular* and *Luffa cylindrica*, which could be distinguished mainly by fruit ridge and several other phenotypic characteristics (Supplementary Figure S1) (Pootakham et al., 2021). In present study, we measured 25 traits related to maturity, flower size, plant growth and fruit (Table 1; Supplementary Figures S2, S3). The results illustrated distinctive phenotypic variations within parents and the derived BC_1 population, and S1174 was found to be an early-maturing but late-daily-flowering cultivar with smaller light-yellow male flower, and longer internode length and fruit, while P93075 performed better in the entire plant growth momentum, fruit peduncle length and old fruit diameter. All of these traits exhibited relatively normal distributions in the BC_1 population, indicating that they are quantitative traits controlling by multiple genes, and the phenotypic distribution pattern of diurnal flowering time was consistent with our previous research of the F_2 population (Wu et al., 2016).

As expected, significant correlations were discovered among many traits within one category, such as that between FFFN and FFRN ($r = 0.83$, $P < 0.01$). While for traits across different categories, the correlation coefficients were relatively low, with the strongest correlation between FFFN, FFRN and TNI ($r = 0.34$, $P < 0.01$) (Supplementary Figure S4; Supplementary Table S5). To our knowledge, this is the largest and most comprehensive dataset of phenotypic analysis in *Luffa* across both cultivars of *Luffa cylindrica* and *Luffa acutangula*.

The first high-density genetic map across interspecies of *Luffa*

In our previous study, a genetic map of *Luffa* using 177 EST-SSR markers was constructed (Wu et al., 2016). However, the

A

Position / Lines	-271	-270	-268	-146	-130	-101	-79	-36
S1174-like (94)	-	-	-	T	TTCCA	TCC	-	G
P93075-like (89)	GGACTTAGGG GCCCTTTTCA	GGGAC TTAGG	TTGGGACTTAGGGGC CCTTTTCACCGGTGTT	T	-	-	AAAAAAG	G

Position / Lines	58	206	6858	7020	11780	11885	13950	13984	14269	14719	14955	15062
S1174-like (94)	C	C	T	A	T	A	C/T	A	T	T	T	A
P93075-like (89)	T	T	C	G	G	C	C	T	A	G	A	T

Position / Lines	15220	15360	15852	15940	16030	16190	16328	16372	16531	16690	30837	30972
S1174-like (94)	G	A	T/A	C	A	G	G	G	T	G	C/T	C
P93075-like (89)	T	C	T	G	T	A	A	C	C	A	C	G

B

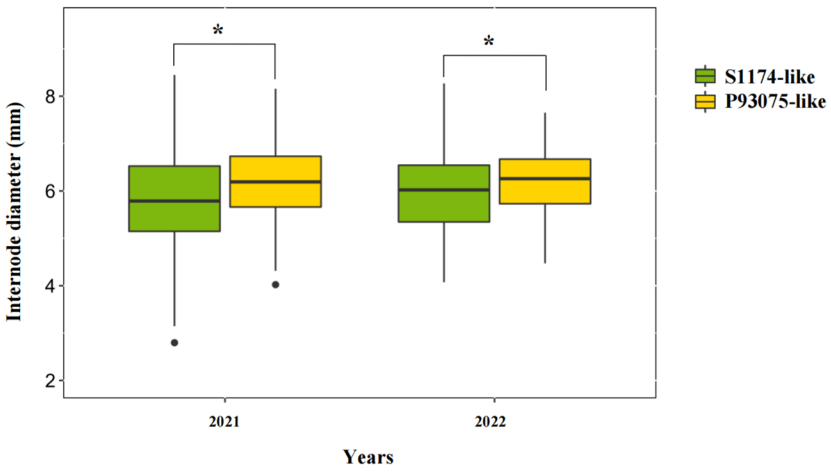


FIGURE 4
Haplotype analysis of *LacCRWN3* among the 183 core inbred line population. **(A)** Distribution of the nonsynonymous mutation single nucleotide polymorphisms (SNPs) in exons and variations in 5′ un-translated region (UTR) for the two haplotypes of S1174-like and P93075-like. The figures in parentheses denote the number of inbred lines belonging to the corresponding haplotypes. The two haplotypes are displayed in green and yellow backgrounds, respectively. **(B)** The average internode diameter differences between the two haplotypes in the 2021 and 2022 seasons. Results indicate mean ± SE of three replicates, $*P < 0.05$.

genome coverage and resolution of the map were relatively poor because of the low polymorphism of SSR makers. With the development of high-throughput sequencing and bioinformatics technology, several *Luffa* genomes have been published successively in recent years (Zhang et al., 2020a; Pootakham et al., 2021). and our team has completed the *de novo* assemblies of S1174 (Unpublished data) and P93075 (Wu et al., 2020). As a result, high-throughput markers may be developed to efficiently improve the genetic map. A *Luffa* genetic map through an F_2 population of 130 individuals crossed by two inbred line parents from *Luffa cylindrica* was constructed based on SLAF-seq, and 3701 of the polymorphic markers were mapped to 13 LGs and the map spanned 1518.56 cM with an average distance of 0.41 cM between adjacent markers (Lou et al., 2020). In this study, we successfully generated a genetic map with higher resolution, and 4,673 SNP markers were mapped to 13 LGs with a total length of 2246.74 cM and the average genetic length of LGs was 359.4 cM (Figure 1; Supplementary Table S4). Most importantly, our map was developed from interspecific crosses between *Luffa acutangula* and *Luffa cylindrica*, in which both phenotypic and genotypic diversity were more abundant. For example, for daily flowering time, approximately a 12-h difference existed between the two cultivars. We believe that our research concerning these traits will assist in breaking interspecific reproductive isolation

FIGURE 5
Phylogenetic and conserved motif analysis of *LacCRWN3*. The protein sequences of the LacCRWN3 family were from *Luffa*, the other eight Cucurbitaceae species, *Arabidopsis*, rice and maize.

and promote the utilization of interspecific heterosis in *Luffa* breeding program in the future. In general, we constructed an interspecific high-quality and dense genetic map for *Luffa*, which will be a valuable genomic resource for dissecting numerous *Luffa* traits.

A genome-wide landscape of genetic controls for multiple agronomic traits in *Luffa*

QTL mapping was performed for the 25 traits linked to maturity, flower size, plant growth and fruit to demonstrate the application of our high-density genetic map. In total, 40 significant QTLs associated with 17 traits were detected across 11 chromosomes except chromosomes 2 and 3. This is so far the largest dataset of genetic controls for agronomic traits in *Luffa* (Figure 2; Supplementary Table S6). Among the 40 identified loci, nearly four-fifths accounted for > 10% of phenotypic variation, indicating that many genes with major effects were responsible for the genetic controls of these traits.

It was the first research to examine the genetic architecture of the majority of these traits in *Luffa*. Zhao et al. (2021) identified a dwarfism gene *LacDWARF1* on chromosome 5 by combined BSA-Seq and comparative genomics analyses, which was 3 Mb away from our QTL of the 9th internode length. In addition, RNA-Seq revealed that plant hormones, cellular process, cell wall, membrane and stress response were significantly related to the dwarf phenotype (Zhao et al., 2021). Consistent with the findings, candidate genes associated with the length of *Luffa* internodes and branches also exhibited enrichment in these biological processes in our study.

Thereafter, we also identified some important homologous genes of key traits reported in other species. MADS transcriptional regulator genes, *AtSEP3* and *AtSOC1*, were involved in the clock output process of the photoperiodic pathway, where *AtSEP3* is a downstream gene of FLOWERING LOCUS T (FT) and *AtSOC1* is a flowering integrator (Hwan Lee et al., 2012; Adal et al., 2021; Fukazawa et al., 2021). Coincidentally, their homolog genes in *Luffa*, *Lac01g012970* and *Lac01g012980*, were located in our QTL of daily flowering time, which may be the most likely candidate genes for further study. Tomato *SUN* was reported to regulate fruit shape by altering cell division patterns (Wu et al., 2011), and its ortholog in watermelon, *Cla011257*, has also been demonstrated to be associated with fruit shape variation and a deletion of 159 bp in its coding region may cause fruit elongation (Dou et al., 2018). Its *Luffa* ortholog, the calmodulin-binding family gene *Lac10g015060*, targeted by *qYFD1*, was also identified to control young fruit shape. This finding demonstrated that our high-density genetic map was effective in elucidating the genetic basis of complex traits in *Luffa*.

Given all of that, the 40 identified QTLs will lay a solid foundation for further research on the genetic and molecular basis of agronomic traits in *Luffa*.

LacCRWN3 regulating the internode diameter variation of *Luffa*

The stem, an important part of plants, performs essential roles in supporting, lodging resistance and nutrient transport to facilitate plant growth and development (Ye et al., 2020). Various molecular mechanisms involved in stem development were confirmed in *Arabidopsis*, including shoot apical meristem

development, cyclin, transcript factor and phytohormone pathways responsible for cell organization (Pierre-Jerome et al., 2018). In cucurbit crops, the research on stem diameter has been mainly focused on genetic analysis and QTL mapping (Li et al., 2008; Rao et al., 2021), and there is no relevant study in *Luffa*. In present study, we identified a significant *qID2* associated with main stem internode diameter on chromosome 9, which explained 11% of the phenotypic variation (Figure 2; Supplementary Table S6). The candidate gene *Lac09g006860* was the only gene harbored in *qID2* and it encoded a CRWN3 protein. CRWN, located in the nucleus, encodes components of putative plant nuclear laminas. Its animal homologous proteins are the major components of the metazoan nuclear lamina (Gruenbaum and Foisner, 2015), which are involved in the maintenance of the chromatin architecture and nuclear shape, gene expression, cell differentiation, and metabolism. Mutations in these proteins cause severe diseases (Eriksson et al., 2003). In recent years, a series of lamina orthologs have also been identified successively from Charophyta to terrestrial plants (Sakamoto, 2020). Typically, *Arabidopsis* has four CRWN proteins, CRWN1 to CRWN4, and they participate in the repair of DNA damage, chromatin organization, transcriptional regulation, and nuclear body formation (Dittmer et al., 2007; Sakamoto, 2020). Their single and double mutants display abnormal nuclear structure and phenotypes (Sakamoto and Takagi, 2013; Wang et al., 2013; Zhao et al., 2016; Wang et al., 2019; Sakamoto, 2020). For example, *crwn1crwn2* showed significantly reduced nuclear size and chromocenter number, CRWN3 played a role in regulating the ABI5 protein level *via* the formation of the degradation body in the nucleus plants, and *crwn1,3* had the severe retarded phenotypes, such as lower germination rates, short roots, small leaves, dwarf seedlings, short siliques, and shriveled seeds, whereas *crwn* triple mutants were usually lethal. Meanwhile, severe genomic DNA damage and upregulated expression levels of DNA damage-responsive genes were also detected in the *crwn* mutants lines (Hirakawa and Matsunaga, 2019). Our candidate gene *Lac09g006860* is the only *CRWN3* in cucurbit crops, and it was evolutionarily close to *AtCRWN3* and *AtCRWN2* (Figure 5). Sequence alignment and qRT-PCR in the two parental lines showed that, except for SNPs in exons, completely different promoter sequences were detected, and the expression level in thick-stem P93075 were distinctively higher than that in thin-stem S1174 (Figures 3A–C). Further natural variation analysis revealed that haplotype P93075-like lines exhibiting significantly thicker stem diameters than those of haplotype S1174-like lines (Figures 4A, B). Moreover, unique ABRE motifs related to ABA responses were found in the P93075 allele, which might be consistent with previous reports that *CRWNs* play key roles in ABA-related seed germination and plant development (Zhao et al., 2016; Wang et al., 2019). The results highlighted that *LacCRWN3* might be a positive regulator of *Luffa* internode diameter, and responsible for the internode diameter divergence between *Luffa acutangular* and *Luffa cylindrica* lines.

Implications for future *Luffa* breeding

Pleiotropic effect is that the QTL responsible for different traits tend to cluster together on a certain region of the chromosome (Verma et al., 2004; Pan et al., 2020b). Our study identified four hotspot regions, and QTLs for strongly correlated traits, including *qFFRN1* and *qFFFN2*, *qOFL1* and *qOFPL1*, were clustered in the same genomic regions, and their increasing effects were all from the *P93075* alleles (Figure 2; Supplementary Table S6). Meanwhile, some QTLs for traits with weak correlations were also found to be clustered together, such as the QTLs for *qIL* and *qYFPL* on chromosome 7, and *qID* and *qOFPL* on chromosome 8 (Figure 2). It is promising that these pleiotropic loci can be utilized to select combinations of various favorable alleles for improving multiple traits through marker-assisted breeding programs of *Luffa* in the future. Moreover, we identified a key candidate gene related to *Luffa* stem development, *LacCRWN3*, whose favorable *P93075-like* allele was significantly positively correlated with stem thickness (Figures 2-4). We believe that the genetic manipulation of *LacCRWN3* will provide an efficient way to improve lodging resistance and nutrient transport in *Luffa*.

In conclusion, we constructed the first high-density genetic map across interspecies of *Luffa* and identified 40 significant genetic loci controlling different agronomic traits. Except for some key orthologous genes retrieved, a unique gene of the Cucurbitaceae family, *LacCRWN3*, was identified responsible for the *Luffa* internode diameter, and its differential expression profiles associated with the sequence variations in promoter regions might be the main causation for stem thickness divergence in the parental lines. Meanwhile, its two haplotypes, the thin-stem S11174-like and thick-stem P93075-like, provide the molecular evidence for explaining the internode diameter difference among inbred lines. Our findings will provide important genetic information and insights into the dissection of complex traits and breeding improvement in *Luffa*.

Data availability statement

The data presented in the study are deposited in the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) repository, accession numbers: CNP0003546 and CNP0003626.

Author contributions

HW and JNL conceived and designed the experiments; HW, GZ, JXL, XL, LD, XZ and HG provided support for field

phenotyping. LL and YG performed the experiments; LL and GZ provided support for bioinformatics analysis; LL analyzed the data and wrote the manuscript; HW and GZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Laboratory of Lingnan Modern Agriculture Project (NZ2021008), the National Natural Science Foundation of China (31872093 and 31902011), the Science and Technology Program of Guangdong Province (2021A1515012500, 2020A0505020006, 2019A050520002, 201904020012, 2018B020202007 and 2021KJ110), the Science and Technology Program of Guangzhou of China (X20210201020 and 202201011453), China Agriculture Research System of MOF and MARA, Agricultural competitive industry discipline team building project of Guangdong Academy of Agricultural Sciences (202103TD and 202114TD), and the Project of Collaborative Innovation Center of GDAAS(XTXM202203).

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.2022.1069618/full#supplementary-material

References

Abdel-Salam, I. M., Awadein, N. E. S., and Ashour, M. (2019). Cytotoxicity of *Luffa cylindrica* (L.) m. roem. extract against circulating cancer stem cells in hepatocellular carcinoma. *J. Ethnopharmacol.* 229, 89–96. doi: 10.1016/j.jep.2018.09.034

Adal, A. M., Binson, E., Remedios, L., and Mahmoud, S. S. (2021). Expression of lavender *AGAMOUS-like* and *SEPALLATA3-like* genes promote early flowering and alter leaf morphology in *Arabidopsis thaliana*. *Planta* 254, 54. doi: 10.1007/s00425-021-03703-3

Anwar, R., Fatima, S., Mattoo, A. K., and Handa, A. K. (2019). Fruit architecture in polyamine-rich tomato germplasm is determined *via* a medley of cell cycle, cell expansion, and fruit shape genes. *Plants (Basel)* 8, 387. doi: 10.3390/plants8100387

Cai, Y., Bartholomew, E. S., Dong, M., Zhai, X., Yin, S., Zhang, Y., et al. (2020). The HD-ZIP IV transcription factor GL2-LIKE regulates male flowering time and fertility in cucumber. *J. Exp. Bot.* 71, 5425–5437. doi: 10.1093/jxb/eraa251

Cao, Y., Diao, Q., Chen, Y., Jin, H., Zhang, Y., and Zhang, H. (2021). Development of KASP markers and identification of a QTL underlying powdery mildew resistance in melon (*Cucumis melo* l.) by bulked segregant analysis and RNA-seq. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.593207

Chen, L., Yun, M., Cao, Z., Liang, Z., Liu, W., Wang, M., et al. (2021). Phenotypic characteristics and transcriptome of cucumber male flower development under heat stress. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.758976

Dittmer, T. A., Stacey, N. J., Sugimoto-Shirasu, K., and Richards, E. J. (2007). *LITTLE NUCLEI* genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell* 19, 2793–2803. doi: 10.1105/tpc.107.053231

Dou, J., Zhao, S., Lu, X., He, N., Zhang, L., Ali, A., et al. (2018). Genetic mapping reveals a candidate gene (*ClFS1*) for fruit shape in watermelon (*Citrullus lanatus* l.). *Theor. Appl. Genet.* 131, 947–958. doi: 10.1007/s00122-018-3050-5

Eriksson, M., Brown, W. T., Gordon, L. B., Glynn, M. W., Singer, J., Scott, L., et al. (2003). Recurrent *de novo* point mutations in lamin a cause Hutchinson–gilford progeria syndrome. *Nature* 423, 293–298. doi: 10.1038/nature01629

Filipowicz, N., and Schaefer, H. (2014). Revisiting *luffa* (Cucurbitaceae) 25 years after c. heiser: species boundaries and application of names tested with plastid and nuclear DNA sequences. *Syst. Bot.* 39, 205–215. doi: 10.1600/036364414X678215

Fukazawa, J., Ohashi, Y., Takahashi, R., Nakai, K., and Takahashi, Y. (2021). DELLA degradation by gibberellin promotes flowering *via* GAF1-TPR-dependent repression of floral repressors in *Arabidopsis*. *Plant Cell* 33, 2258–2272. doi: 10.1093/plcell/koab102

Gowtham, K. N. P., Kuppast, I. J., and Mankani, K. L. (2012). A review on *Luffa acutangula*. *Int. J. Pharm. World Res.* 3.

Gruenbaum, Y., and Foisner, R. (2015). Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. *Annu. Rev. Biochem.* 84, 131–164. doi: 10.1146/annurev-biochem-060614-034115

Guo, Y., Gao, M., Liang, X., Xu, M., Liu, X., Zhang, Y., et al. (2020). Quantitative trait loci for seed size variation in cucurbits - a review. *Front. Plant Sci.* 11, 304. doi: 10.3389/fpls.2020.00304

Hirakawa, T., and Matsunaga, S. (2019). Characterization of DNA repair foci in root cells of arabidopsis in response to DNA damage. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00990

Hwan Lee, J., Joon Kim, J., and Ahn, J. H. (2012). Role of *SEPALLATA3* (*SEP3*) as a downstream gene of *miR156-SPL3-FT* circuitry in ambient temperature-responsive flowering. *Plant Signal. Behav.* 7, 1151–1154. doi: 10.4161/psb.21366

Jiang, L., Yan, S., Yang, W., Li, Y., Xia, M., Chen, Z., et al. (2015). Transcriptomic analysis reveals the roles of microtubule-related genes and transcription factors in fruit length regulation in cucumber (*Cucumis sativus* l.). *Sci. Rep.* 5, 8031. doi: 10.1038/srep08031

Kircher, S., Gil, P., Kozma-Bognár, L., Fejes, E., Speth, V., Husselstein-Muller, T., et al. (2002). Nucleocytoplasmic partitioning of the plant photoreceptors phytochrome a, b, c, d, and e is regulated differentially by light and exhibits a diurnal rhythm. *Plant Cell* 14, 1541–1555. doi: 10.1105/tpc.00115

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

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

Li, X., Yuan, X., Su, J., Pan, J., Deng, S., Wang, G., et al. (2008). Detecting QTLs for plant architecture traits in cucumber (*Cucumis sativus* l.). *Breed. Sci.* 58, 453–460. doi: 10.1270/jsbbs.58.453

Lou, L., Su, X., Liu, X., and Liu, Z. (2020). Construction of a high-density genetic linkage map and identification of gene controlling resistance to cucumber mosaic

virus in *Luffa cylindrica* (L.) roem. based on specific length amplified fragment sequencing. *Mol. Biol. Rep.* 47, 5831–5841. doi: 10.1007/s11033-020-05652-8

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

Mulekar, J. J., and Huq, E. (2012). Does CK2 affect flowering time by modulating the autonomous pathway in *Arabidopsis*? *Plant Signal. Behav.* 7, 292–294. doi: 10.4161/psb.18883

Pan, Y., Wang, Y., McGregor, C., Liu, S., Luan, F., Gao, M., et al. (2020a). Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. *Theor. Appl. Genet.* 133, 1–21. doi: 10.1007/s00122-019-03481-3

Pan, Y., Wen, C., Han, Y., Wang, Y., Li, Y., Li, S., et al. (2020b). QTL for horticulturally important traits associated with pleiotropic andromonoecy and carpel number loci, and a paracentric inversion in cucumber. *Theor. Appl. Genet.* 133, 2271–2290. doi: 10.1007/s00122-020-03596-y

Papanicolaou, G. C., Psarra, E., and Anastasiou, D. (2015). Manufacturing and mechanical response optimization of epoxy resin/*Luffa cylindrica* composite. *J. Appl. Polym. Sci.* 132, 41992. doi: 10.1002/app.41992

Pierre-Jerome, E., Drapek, C., and Benfey, P. N. (2018). Regulation of division and differentiation of plant stem cells. *Annu. Rev. Cell. Dev. Biol.* 34, 289–310. doi: 10.1146/annurev-cellbio-100617-062459

Pootakham, W., Sonthirod, C., Naktang, C., Nawae, W., Yoocha, T., Kongkachana, W., et al. (2021). *De novo* assemblies of *Luffa acutangula* and *Luffa cylindrica* genomes reveal an expansion associated with substantial accumulation of transposable elements. *Mol. Ecol. Resour.* 21, 212–225. doi: 10.1111/1755-0998.13240

Rabei, S., Rizk, R. M., and Khedr, A.-H. A. (2013). Keys for and morphological character variation in some Egyptian cultivars of cucurbitaceae. *Genet. Resour. Crop Evol.* 60, 1353–1364. doi: 10.1007/s10722-012-9924-5

Rao, P. G., Behera, T. K., Gaikwad, A. B., Munshi, A. D., Srivastava, A., Boopalakrishnan, G., et al. (2021). Genetic analysis and QTL mapping of yield and fruit traits in bitter gourd (*Momordica charantia* l.). *Sci. Rep.* 11, 4109. doi: 10.1038/s41598-021-83548-8

Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. doi: 10.1093/bioinformatics/btx494

Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* 176, 836–850. doi: 10.1104/pp.17.01290

Sakamoto, Y. (2020). Nuclear lamina CRWN proteins regulate chromatin organization, gene expression, and nuclear body formation in plants. *J. Plant Res.* 133, 457–462. doi: 10.1007/s10265-020-01184-1

Sakamoto, Y., and Takagi, S. (2013). LITTLE NUCLEI 1 and 4 regulate nuclear morphology in *Arabidopsis thaliana*. *Plant Cell Physiol.* 54, 622–633. doi: 10.1093/pcp/pct031

Shan, N., Xiang, Z., Sun, J., Zhu, Q., Xiao, Y., Wang, P., et al. (2021). Genome-wide analysis of valine-glutamine motif-containing proteins related to abiotic stress response in cucumber (*Cucumis sativus* l.). *BMC Plant Biol.* 21, 492. doi: 10.1186/s12870-021-03242-9

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8, e58700. doi: 10.1371/journal.pone.0058700

Sun, Y., Zhang, H., Dong, W., He, S., Qiao, S., Qi, X., et al. (2022). Integrated analysis of the transcriptome, sRNAome, and degradome reveals the network regulating fruit skin coloration in sponge gourd (*Luffa cylindrica*). *Sci. Rep.* 12, 3338. doi: 10.1038/s41598-022-07431-w

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

Tyagi, R., Sharma, V., Sureja, A. K., Das Munshi, A., Arya, L., Saha, D., et al. (2020). Genetic diversity and population structure detection in sponge gourd (*Luffa cylindrica*) using ISSR, SCoT and morphological markers. *Physiol. Mol. Biol. Plants* 26, 119–131. doi: 10.1007/s12298-019-00723-y

Van Ooijen, J. W. (2009). *MapQTL 6.0. software for the mapping of quantitative trait loci in experimental populations of diploid species*. Ed. B. V. Kyazma (Netherlands: Wageningen).

Verma, V., Foulkes, M. J., Worland, A. J., Sylvesterbradley, R., Caligari, P. D. S., and Snape, J. W. (2004). Mapping quantitative trait loci for flag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. *Euphytica* 135, 255–263. doi: 10.1023/B:EUPH.0000013255.31618.14

Wang, S., Basten, C., and Zeng, Z. (2012). *Windows QTL cartographer 2.5* (North Carolina State University, USA: Department of Statistics).

Wang, Y., Bo, K., Gu, X., Pan, J., Li, Y., Chen, J., et al. (2020). Molecularly tagged genes and quantitative trait loci in cucumber with recommendations for QTL nomenclature. *Hortic. Res.* 7, 3. doi: 10.1038/s41438-019-0226-3

Wang, H., Dittmer, T. A., and Richards, E. J. (2013). *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol.* 13, 200. doi: 10.1186/1471-2229-13-200

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi: 10.1093/nar/gkq603

Wang, H., Li, W., Qin, Y., Pan, Y., Wang, X., Weng, Y., et al. (2017a). The cytochrome P450 gene *CsCYP85A1* is a putative candidate for *super compact-1* (*Scp-1*) plant architecture mutation in cucumber (*Cucumis sativus* l.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00266

Wang, Q., Liu, S., Lu, C., La, Y., Dai, J., Ma, H., et al. (2019). Roles of CRWN-family proteins in protecting genomic DNA against oxidative damage. *J. Plant Physiol.* 233, 20–30. doi: 10.1016/j.jplph.2018.12.005

Wang, Y., Zhao, J., Lu, W., and Deng, D. (2017b). Gibberellin in plant height control: old player, new story. *Plant Cell Rep.* 36, 391–398. doi: 10.1007/s00299-017-2104-5

Wen, H., Pan, J., Chen, Y., Chen, G., Du, H., Zhang, L., et al. (2021). TERMINAL FLOWER 1 and TERMINAL FLOWER 1d respond to temperature and photoperiod signals to inhibit determinate growth in cucumber. *Plant Cell Environ.* 44, 2580–2592. doi: 10.1111/pce.14075

Wu, H., Gong, H., Liu, P., He, X. L., Luo, S., Zheng, X., et al. (2014). Large-Scale development of EST-SSR markers in sponge gourd via transcriptome sequencing. *Mol. Breed.* 34, 1903–1915. doi: 10.1007/s11032-014-0148-6

Wu, H., He, X., Gong, H., Luo, S., Li, M., Chen, J., et al. (2016). Genetic linkage map construction and QTL analysis of two interspecific reproductive isolation traits in sponge gourd. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00980

Wu, H., He, X., Gong, H., Luo, S., Li, M., Chen, J., et al. (2020). A high-quality sponge gourd (*Luffa cylindrica*) genome. *Hortic. Res.* 7, 128. doi: 10.1038/s41438-020-00350-9

Wu, S., Xiao, H., Cabrera, A., Meulia, T., and van der Knaap, E. (2011). *SUN* regulates vegetative and reproductive organ shape by changing cell division patterns. *Plant Physiol.* 157, 1175–1186. doi: 10.1104/pp.111.181065

Xin, T., Zhang, Z., Li, S., Zhang, S., Li, Q., Zhang, Z. H., et al. (2019). Genetic regulation of ethylene dosage for cucumber fruit elongation. *Plant Cell* 31, 1063–1076. doi: 10.1105/tpc.18.00957

Yang, Y., Dong, S., Miao, H., Liu, X., Dai, Z., Li, X., et al. (2022). Genome-wide association studies reveal candidate genes related to stem diameter in cucumber (*Cucumis sativus* l.). *Genes (Basel)* 13, 1095. doi: 10.3390/genes13061095

Ye, J., Tian, R., Meng, X., Tao, P., Li, C., Liu, G., et al. (2020). Tomato *SD1*, encoding a kinase-interacting protein, is a major locus controlling stem development. *J. Exp. Bot.* 71, 3575–3587. doi: 10.1093/jxb/eraa144

Zhang, S., Hu, J., Zhang, C. F., Guan, Y. J., and Zhang, Y. (2007). Genetic analysis of fruit shape traits at different maturation stages in sponge gourd. *J. Zhejiang Univ. Sci. B* 8, 338–344. doi: 10.1631/jzus.2007.B0338

Zhang, T., Ren, X., Zhang, Z., Ming, Y., Yang, Z., Hu, J., et al. (2020a). Long-read sequencing and *de novo* assembly of the *Luffa cylindrica* (L.) roem. *genome. Mol. Ecol. Resour.* 20, 511–519. doi: 10.1111/1755-0998.13129

Zhang, Z., Wang, B., Wang, S., Lin, T., Yang, L., Zhao, Z., et al. (2020b). Genome-wide target mapping shows histone deacetylase complex1 regulates cell proliferation in cucumber fruit. *Plant Physiol.* 182, 167–184. doi: 10.1104/pp.19.00532

Zhao, W., Guan, C., Feng, J., Liang, Y., Zhan, N., Zuo, J., et al. (2016). The *Arabidopsis CROWDED NUCLEI* genes regulate seed germination by modulating degradation of ABI5 protein. *J. Integr. Plant Biol.* 58, 669–678. doi: 10.1111/jipb.1244

Zhao, J., Jiang, L., Che, G., Pan, Y., Li, Y., Hou, Y., et al. (2019). A functional allele of *CsFUL1* regulates fruit length through repressing *CsSUP* and inhibiting auxin transport in cucumber. *Plant Cell* 31, 1289–1307. doi: 10.1105/tpc.18.00905

Zhao, G., Luo, C., Luo, J., Li, J., Gong, H., Zheng, X., et al. (2021). A mutation in *LacDWARF1* results in a GA-deficient dwarf phenotype in sponge gourd (*Luffa acutangula*). *Theor. Appl. Genet.* 134, 3443–3457. doi: 10.1007/s00122-021-03938-4

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 November 2022
DOI 10.3389/fpls.2022.1069750

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Libo Han,
Fujian Agriculture and Forestry University, China
Xiaoqin Liu,
Peking University, China

*CORRESPONDENCE
Yan Yang
yziqi@126.com

†These authors have contributed equally to this work

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

RECEIVED 14 October 2022
ACCEPTED 25 October 2022
PUBLISHED 22 November 2022

CITATION
Yu R, Niu Y, Wang X, Yang K, Han X, Liu Z, Qi Z and Yang Y (2022) Construction of a density mutant collection in bitter gourd *via* new germplasms innovation and gene functional study.
Front. Plant Sci. 13:1069750.
doi: 10.3389/fpls.2022.1069750

Construction of a density mutant collection in bitter gourd *via* new germplasms innovation and gene functional study

Renbo Yu[1,2,3†], Yu Niu[1,2,3†], Xiaoyi Wang[4†], Kaili Yang[1,2], Xu Han[1,2,3], Zhaohua Liu[1,3], Zhiqiang Qi[1,2,3] and Yan Yang[1,2,3*]

[1]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China, [2]Hainan Yazhou Bay Seed Lab, Hainan, China, [3]Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Hainan, China, [4]Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

Although a few studies have elucidated the creation of bitter gourd mutants, the suitable concentration and duration of ethyl methanesulfonate (EMS) mutagenesis have not been determined. In this study, mutant collection was conducted to create new germplasms and widen genetic diversity. By employing the seeds of the inbred line Y52 as the mutagenic material, EMS as the mutagen, and the suitable mutagenic conditions for bitter gourd seeds (EMS concentration 0.2%, mutagenic time 10 h), we mutated 10,000 seeds and acquired 3223 independent M_1 lines. For the randomly selected 1000 M_2 lines, 199 M_2 lines with visible phenotypes were found, and 167 M_2 lines were mutants of fruit shape, size, and tubercles. Furthermore, fourteen dwarf, eleven leaf color, five leaf shape, and eight meristem defect mutants were discovered in this mutant collection. In addition, three lines of 1253, 2284, and 3269 represented recessive mutants crossed with Y52. Furthermore, the yellow leaf lines of 2284 and 3269 were not mutated at the same gene locus. This study constructed a mutant collection through innovative new germplasms and provided valuable resources for bitter gourd breeding and functional gene research.

KEYWORDS

bitter gourd, EMS, mutant collection, phenotype, germplasm

Introduction

Bitter gourd (*Momordica charantia* L., 2n = 22), also named balsam pear, bitter melon, bitter cucumber, or African cucumber, belongs to the family *Cucurbitaceae*, and is cultivated in Asia and Africa. As an important vegetable, it has many other pharmacological benefits for human immunodeficiency virus (HIV) infection and

diabetes and plays a vital role in our life (Behera et al., 2010). Owing to land shortages and the worsening environment, breeding is focused on promoting crop yield. Bitter gourd is facing the problem of using less land to obtain higher yields, and biotechnology can solve this problem. For a long time, the basic research of bitter gourd was concentrated on establishing recombinant inbred lines and constructing genetic linkage maps using molecular markers (Cui et al., 2017; Cui et al., 2018; Rao et al., 2018; Kaur et al., 2021; Yang et al., 2022; Zhong et al., 2022).

Creating mutant materials to conduct the functional study of related genes in genetic research is necessary when the ratio of spontaneous variation is very low. Consequently, an efficient method for creating mutants is urgently required. Recently, gene editing technology with the CRISPR-Cas system has been widely used in rice, wheat, and tomato (Li et al., 2018; Lin et al., 2020; Wang et al., 2022), but it is coupled with an efficient transformation rate. Nevertheless, the transformation system has not been established in bitter gourd, resulting in the impossibility of gene editing in bitter gourd.

In bitter gourd, the functional study of genetics is considerably behind other *Cucurbitaceae* crops like watermelon and cucumber because of the lack of mutants (Deng et al., 2022; Zhang et al., 2022). A long-recognized large-scale method for mutant acquisition is mutagenization using chemical or physical mutagens. Physical mutagens such as fast neutrons, gamma rays, and x-rays mainly cause nucleotide deletion, and their mutation frequency is low (Li et al., 2017). As a chemical mutagen, EMS mutagenesis was successfully used in crop breeding and functional genomics research. EMS mutagenesis technology is feasible in bitter gourd because it is appropriate to most plants rather than their transformation rate (Lian et al., 2020).

EMS is an effective mutagen that can introduce random base pair changes (mainly G:C/AT) in the genome after replication, bringing less damage to plants (Greene et al., 2003). For example, in *Arabidopsis thaliana*, approximately 700 mutations in each EMS mutant line and 50000 M_1 lines are sufficient to make a 95% chance of mutation in any G:C base pair in the genome (Jander et al., 2003). Besides, the mutants at M_2 or more advanced generations of recessive homozygous are crossed with the world-type and generate second filial generation (F_2) progeny that can be easily discovered by the mutated genes using MutMap (Abe et al., 2012; Qu and Qin, 2014).

Because EMS mutagenesis is easy to control, exhibits a high mutation rate, and causes less harm to plants, it was widely used in rice (Henry et al., 2014), maize (Lu et al., 2018), soybean (Li et al., 2017), pepper (Arisha et al., 2015), cabbage (Sun et al., 2022), watermelon (Deng et al., 2022), cucumber (Zhang et al., 2022) and *Gossypium hirsutum* L (Lian et al., 2020). EMS concentrations differ from one crop to another, and a suitable mutagen concentration can be used to achieve the desired effect. Generally, the optimal lethality among the M_1 generation is approximately 50% (Arisha et al., 2015; Lian et al., 2020; Dutta et al., 2021). However, the optimal and specific protocol of EMS treatment conditions has not been determined because of limited studies on EMS treatment experiments in bitter gourd. Therefore, there is an urgent need to conduct an EMS mutagenesis study and determine the suitable concentration and duration time of EMS for bitter gourd seeds.

There are few studies on the creation of bitter gourd mutants. Some mutants related to vine length, fertility, and nutrient contents were generated by gamma rays (Co_{60} source) on a small scale (Dutta et al., 2021); however, they did not significantly improve functional genomic research and breeding. Therefore, this study focused on establishing an optimal EMS technique and guidance for other scientists in bitter gourd research. Furthermore, the mutant collection successfully created by EMS mutagenesis will contribute new germplasms to bitter gourd breeding and accelerate the process of gene functional analysis.

Materials and methods

Plant material

The cultivated inbred line Y52 was used in this study, and artificial self-pollinated seeds were supplied by the Institute of Tropical Crop Genetic Resources in Hainan province, China.

EMS experiment

Ten thousand seeds with a small opening at the seed coat of Y52 were soaked in ddH_2O overnight and then embedded in 0.2% EMS at room temperature with gentle shaking for 10 h. The treated seeds were detoxified with 0.1 mol/L sodium thiosulfate five times. Afterward, the seeds were washed 15 times with ddH_2O for 5 min each time. Finally, the seeds were wrapped with a wet cloth and transferred to an incubator at 37 °C in a dark condition until germination.

The germinated seeds were planted into small pots 3–5 days later and placed in a greenhouse until 4–6 true leaves appeared. Then, the seedlings were sown into the soil in an open field. By artificial self-pollination, 3223 M_1 plants generated M_2 mature seeds, and 10–60 M_2 seeds were harvested, labeled, and saved. In addition, 10–20 M_2 seeds from each individual M_1 plant were randomly selected to sow into the soil, and their morphological phenotype was detected.

Results

EMS mutagen experiment condition

Because high concentrations of EMS and prolonged treatment can damage seeds and cause germination problems,

we first detected the suitable conditions for bitter gourd mutagenesis. To increase the efficiency of mutagenesis, we created an opening at the embryo of each seed coat. Then, the preliminary assay of EMS treatment was completed, and we counted the germination seeds, and statistical analysis was done 15 days later (Table 1).

In our preliminary assay, the control group seeds' germination rate was more than 90%. However, the germination rate with 0.5% EMS treatment in 4 h or 10 h were all lower than 50%, indicating this dose of EMS was not suitable for bitter gourd seeds. In contrast, when the seeds were treated with 0.2% EMS for 10 h, the germination rate was around 50%–60%, which was close to the half-lethal dose, indicating 0.2% EMS for 10 h treatment is the right condition for bitter gourd seeds. Besides, we also used 0.2% EMS for 4 h treatment and found that the mutagenesis effect is poor for bitter gourd.

M_1 Population

Y52 is a cultivated species with a length of 30–40 cm and a weight of about 0.5 kg. Therefore, it was selected as the basic material for mutagenesis. Nearly 10,000 seeds were treated with 0.2% EMS for 10 h, and 5319 seeds were germinated, then all the germinated seedlings were planted in the soil (Table 2). After one month, we found that 4356 plants survived and 783 plants died, which may have been caused by the injury of the EMS mutagen.

M_2 Population

Bitter gourd is a monoecious plant with staminate and pistillate flowers born on different internodes. Self-pollination is required for bitter gourd mutant propagation. However, some plants failed to pollinate because female and male flowers did not bloom simultaneously, resulting in no seeds. Finally, only 3223 M_1 plants were successfully harvested (Table 2), and we constructed a bitter gourd mutant collection containing 3223 mutants.

Because bitter gourd needs artificial pollination and has a large workload, we randomly selected 1000 mutants for planting, 10–20 seeds for each, and approximately 10,000–20,000 seedlings were planted. As a result, the phenotypic variation of M_2 generation was very abundant, approximately 19.9% (Table 3), including plant height, leaf color, leaf shape, fruit shape, and meristem defect.

TABLE 1 The germination rate of different EMS concentration treatment.

EMS concentration and duration time	Total number of seeds	The number of germination seeds	Germination rate
0.2% EMS-4h	200	136	68%
	200	141	70.5%
0.2% EMS-10h	200	117	58.5%
	200	106	53%
0.5% EMS-4h	200	80	40%
	200	81	40.5%
0.5% EMS-10h	200	36	18%
	200	33	16.5%
0.1 % Tween20-10h	200	186	93%
	200	193	96.5%

Germination rate: germination seeds / Total number of seeds.

TABLE 2 The germination and survival rate of the M1 generation.

Treatments	0.1 % Tween20	EMS
Total number of seeds	200	10000
Number of germination seeds	191	5319
Number of survival seedlings	186	4536
Fruitful	186	3223
Seed germination rate	95.50%	53.19%
Survival rate	97.38%	85.28%
Fruitful rate	100.0%	71.05%

Germination rate: Number of germination seeds / Total number of seeds, Survival rate: Number of survival seedlings / Number of germination seeds, Fruitful rate: Fruitful / Number of survival seedlings.

Dwarf phenotype

Approximately fourteen mutants were not higher than 60 cm or had short internodes, leading to low plant height (Figure 1). Among fourteen mutants, nine lines of 454, 931, 1313, 1372, 2078, 2981, 3070, 3080, and 3114 had arrested growth and were no more than 30 cm tall. Moreover, there were 3–4 plants with dwarf phenotype in lines 931, 2981, and 3070 with a segregation rate close to 4:1 (Supplemental Table 1), consistent with Mendel's classical genetic law. In contrast, there were only one or two mutants in other lines, which was inconsistent with Mendelian inheritance law. The possible reason is that the number of family plants is insufficient and does not have a statistical effect. Additionally,

TABLE 3 The mutants of M2 generation.

Mutant phenotype	Number of families	%
Dwarf	14	1.4%
Leaf color	11	1.1%
Leaf shape	5	0.5%
Meristem defect	8	0.8%
Fruit shape	15+152=167	15.8%
In all	205-6=199	19.9%

15+152, "15" indicate 15 lines in Figure 5, "+152" indicate another 152 lines in Supplemental Table 5. "-6" means 6 lines of 681, 1084, 1313, 1372, 1479 and 2078 analyzed twice in this study.

FIGURE 1
Different dwarf mutants in the M_2 generation compared to Y52. Scale bar in red color = 5 cm.

lines 681, 942, and 2194 are not only nanoid but also have small leaves, and lines 1833 and 3210 have short internodes (Figure 1).

Leaf color

Besides dwarf, leaf color is another easily observed trait. A total of eleven lines had color changes owing to chlorophyll synthesis defects (Abe et al., 2012), mainly leaf yellowing and leaf albinism (Figure 2). The leaves of the whole plant line 239, 3135, 3186, and 3206 were light yellow, and five lines of 1091, 1668, 2078, 3217, and 3269 had yellow leaves at the top of the plants while the leaves of line 2824 near the ground were yellow. Furthermore, we also found an albino mutant line 2401 with partial leaves is albinism. Of all these mutant lines, 1668, 2401, 2824, and 3269 had a segregation of 4:1 (Supplemental Table 2).

Leaf shape

We selected five mutants with prominent variations in leaf shape, namely 372, 681, 1253, 2002, and 3064 (Figure 3). The leaf blades of lines 372 and 2002 were upward curling, and 2002 is more serious. Line 681 has a small leaf size with some blades bulging upward. Unlike 681, the leaf edge of 3064 is blunt. The mutant of 1253 has severe defect with withered leaves, but they can grow and bloom with fewer female flowers. Unfortunately, the five mutants have the same sterility problem and can only be conservated by heterozygotes which increases the difficulty of gene localization. Moreover, the segregation ratio of 372, 1253, and 3064 are close to 4:1 except the line 681and 2002 (Supplemental Table 3).

Shoot apical meristem defect

The shoot apical meristem defect directly affected the yield of bitter gourd. Eight lines with shoot apical meristem defects were obtained by mutant collection screening (Figure 4). Three lines of 345, 1372, and 3211 have flat and wide shoot apical meristem, and the apical meristem of lines 1084, 1479, 1519, and 2837 develops into petiole, while the apical meristem of line 1319 was dead or withered. In addition, the segregation of all the lines was inconsistent with the Mendelian genetic law due to insufficient plants (Supplemental Table 4).

Fruit shape mutants

In our study, the variation of fruit shape was the most abundant phenotype, including fruit length, fruit transverse diameter, the pattern of fruit bibbing and warts, the shape of the blossom end, and the shape of the base of the fruit. From the 1000 M_2 lines, we screened 167 mutants with various fruit shapes (Supplemental Table 5). Among the 167 mutants, eighteen mutants (Figure 5A) with a nearly round phenotype at the blossom end and thirty-eight mutants (Figure 5B) with a truncated phenotype at the base of the fruit were found compared with Y52.

Besides, compared with Y52 (Figure 6A) there were three kinds of mutation related to the pattern of fruit bibbing and warts: flat, grain, and stripe warts (Figures 6B–D) of three,

10.3389/fpls.2022.1069750

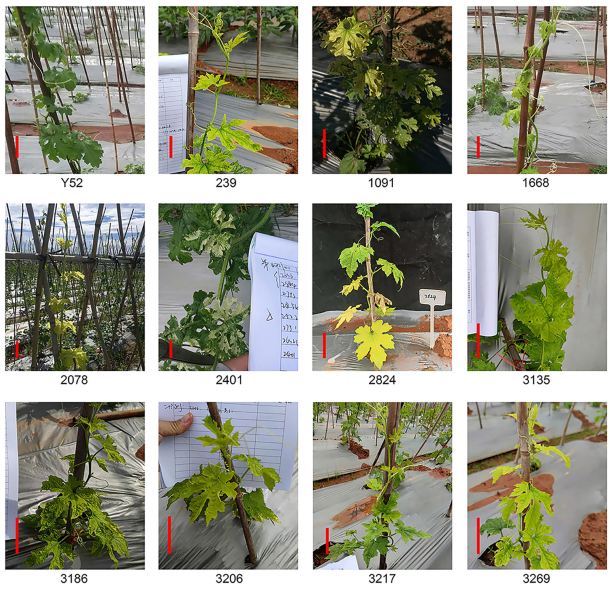


FIGURE 2
Leaf color mutants of M_2 plants compared to Y52. Scale bar in red color = 5 cm.

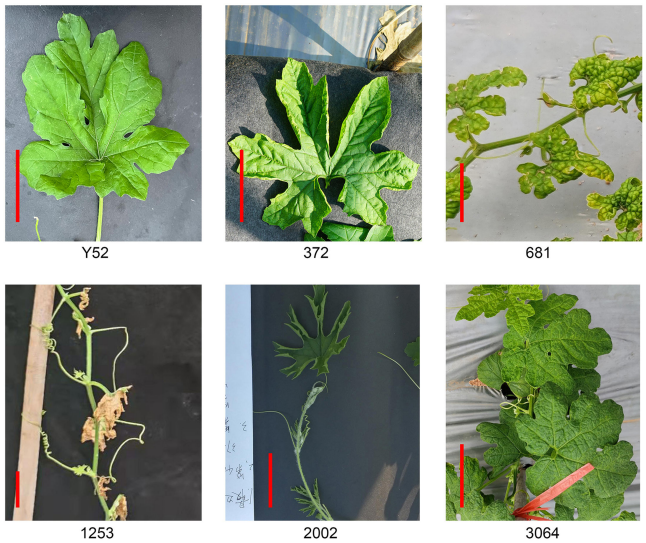


FIGURE 3
Mutants with a variety of leaf shapes compared to Y52. Scale bar in red color = 5 cm.

twenty-eight, and nine mutants, respectively. Second, we found fifteen slender fruit mutants with a small transverse diameter of 3–6 cm (Figure 6E). Additionally, there were twenty-six mutants with fruit lengths less than 20 cm, while the fruit length of Y52 was between 30–40 cm (Figure 6F). Bending fruit is another prominent phenotype, and twenty-seven mutants were found (Figure 6G).

Mutant analysis

The mutant of 1253 has seventeen plants in the M_2 generation and five plants with withered leaves, and the segregation ratio was close to 4:1, which means that it is a recessive mutation. To check the recessive or dominant mutation, we hybridized 1253 with Y52 and found that the

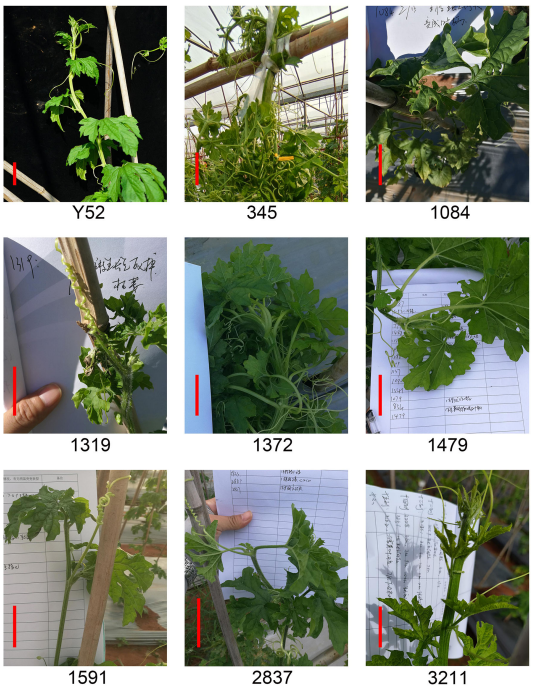


FIGURE 4
Meristem defect mutants of M_2 generation compared to Y52. Scale bar in red color = 5 cm.

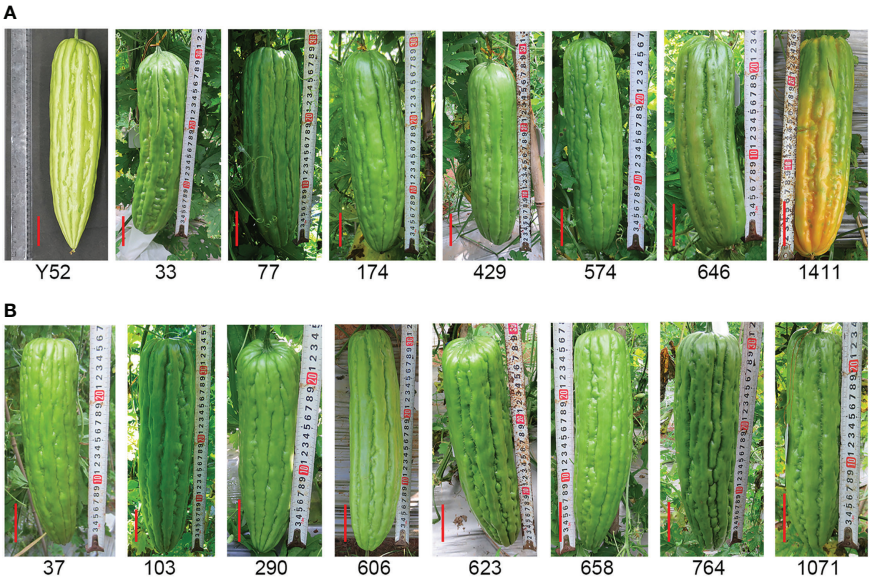


FIGURE 5
Mutation for fruit shape. **(A)** Mutants with nearly round phenotype at the blossom end. **(B)** Mutants with a truncated phenotype at the base of the fruit compared with Y52. Scale bar in red color = 5 cm.

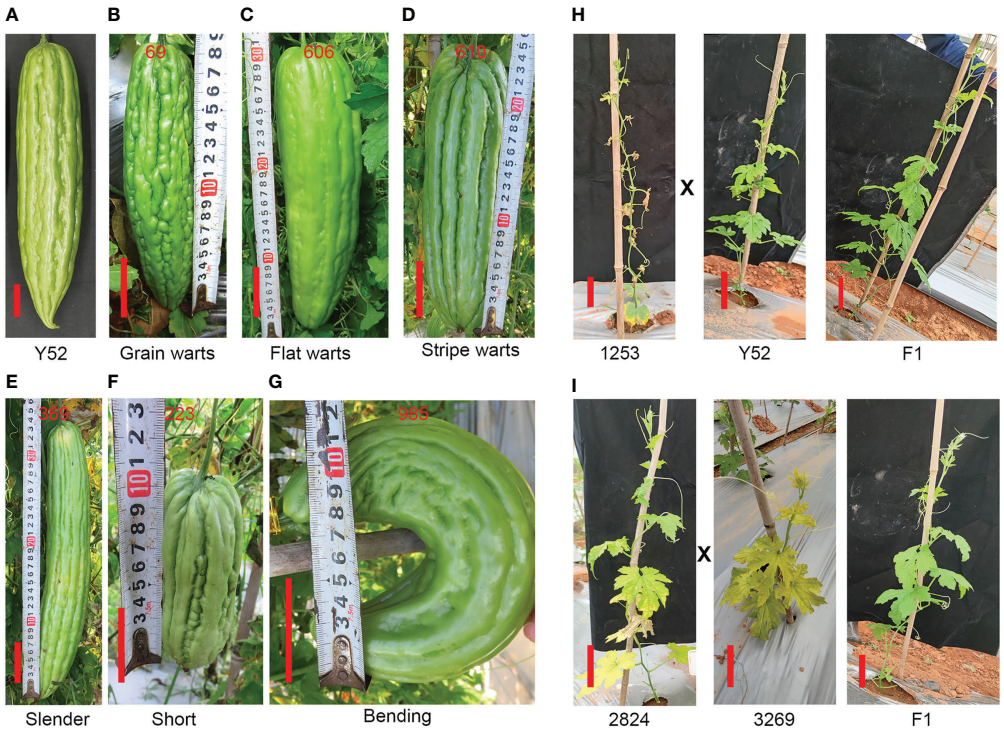


FIGURE 6
Different kinds of phenotype of fruit **(A–G)** and the phenotype of the F_1 generation **(H, I)**. **(A)** Fruit of Y52. **(B–D)** The phenotype of fruit with grain, flat and stripe warts. **(E–G)** Slender, short and bending fruit. The phenotype of the F_1 generation generated by Y52 crossed with 1253 **(H)**, and 2824 crossed with 3269 **(I)** compared to Y52, respectively. Scale bar in red color = 5 cm.

phenotype of the F_1 generation was similar to Y52, which indicated that 1253 was a recessive mutant (Figure 6H). Furthermore, we selected two mutants of 2824 (segregation ratio 4:1) and 3269 (segregation ratio 4:1) with yellow leaves to establish whether the two mutants mutated at the same gene. We hybridized 2824 with 3269 and found that the phenotype of the F_1 generation was similar to Y52, indicating that the two mutants had different gene mutations (Figure 6I).

Specific mutants

Three mutants of 620, 681, and 1152 were selected as specific mutants. Some shoots, leaves, and fruits of 620 were albino, though the seed coat was black; the next generation of seedlings was albino and did not survive (Figure 7A). The mutant of 681 has two plants with a dwarf phenotype and thumb-sized fruits without seeds (Figure 7B), and it was used as a male parent and hybridized with Y52 to keep it. The mutant of 1152 was another interesting mutant with a yellow ovary and premature senescence of leaves attached to charcoal maggot disease (Figure 7C). Unfortunately, the segregation rate of the three mutants was inconsistent with Mendel's classical genetic law, and the difficulty of gene mapping increased.

Discussion

Unlike *Arabidopsis thaliana*, rice, wheat, and soybean, the bitter gourd seed covered with a hard coat hindered the full contact between seed and mutagen, leading to decreased mutagenesis efficiency. To solve the problem and improve the mutagenesis efficiency, we used tweezers to clamp a small opening at the seed coat to enable the EMS mutagen to come into full contact with the seeds. Simultaneously, we added 0.1% Tween 20 to enable the seeds to encounter the mutagen evenly and easily clean. In this study, different EMS concentrations and mutagenesis times were evaluated. The germination rate of bitter gourd seeds treated with 0.2% EMS for 10 h was the closest to 50% (Arisha et al., 2015; Lian et al., 2020). Therefore, this condition was selected for subsequent experiments.

Because most of the M_1 mutants acquired by the EMS mutagenesis were heterozygotes, abundant phenotypic variation was observed in the M_2 generation. Similar to our results, leaf color mutation was easy to find, and we had already discovered eleven mutants with yellow or yellow-green leaves, and the chlorophyll synthesis deficiency mainly caused the yellow leaf mutants found in previous studies (Abe et al., 2012). Although the leaf color mutants survived but had

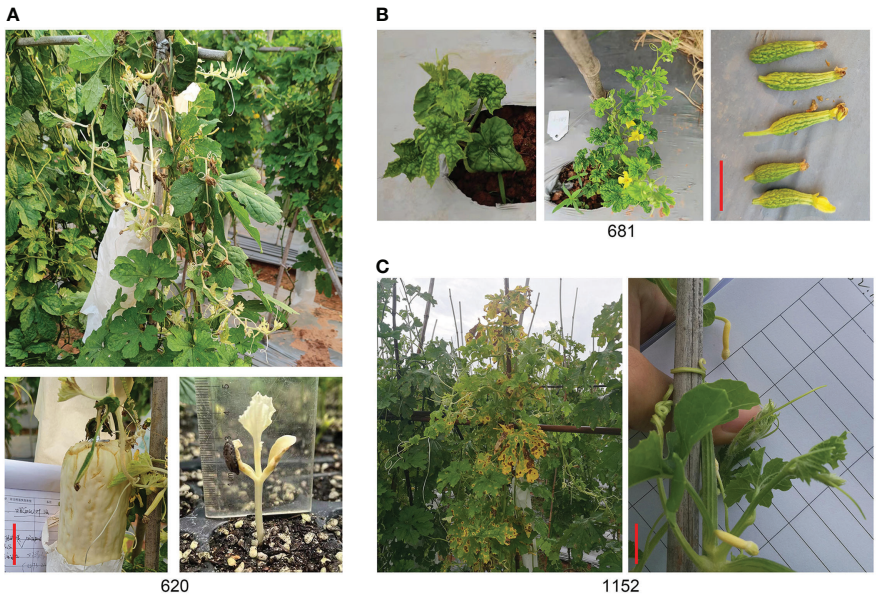


FIGURE 7
Specific mutants. **(A)** Albino mutant of line 620. **(B)** Thumb fruit of line 681. **(C)** Yellow fruit of line 1152. Scale bar in red color = 5 cm.

smaller plants and fruits. Identifying mutation sites that lead to chlorophyll synthesis deficiency is crucial to breeders. Consistent with our results, many gene mutations can cause leaf color deficiency (Ma et al., 2017; Rong et al., 2019), the F_1 generation produced by 2824 and 3269 did not have yellow leaves (Figure 6I), and they have different gene mutations.

Leaf shape, including size, margin, and rolling. Two specific mutants, 372 and 2002, with curly leaves were found in the M_2 generation. Moderate leaf rolling can help to construct the ideal plant architecture and promote photosynthetic efficiency (Rong et al., 2019). Many genes that control leaf rolling in rice have been cloned, such as *CFL2*, *SRL1*, *ZHD1*, *REL1/2*, and *SLL2* (Zhang et al., 2020). In the family of *Cucurbitaceae*, the *CsPHB* gene was regulated by *miRNA165/166* to control leaf curly in cucumber but was never reported in bitter gourd (Rong et al., 2019). Thus, the two leaf curly mutants were valuable germplasms in bitter gourd and will decipher the gene function of the interest phenotype.

Dwarf mutants are crucial to understanding the regulatory mechanisms for plant height, development, and productivity (Bae et al., 2021) because they can be planted at high density and are resistant to lodging (Zhu et al., 2019; Sun et al., 2020). In addition, a class of plant hormones such as GA, BR, and cytokinin were reported to regulate cell elongation and division that could cause the dwarf phenotype when the level of hormones was abnormal, and many dwarf mutants have been found in various crops such as rice *sd1*, wheat *Rht-B1b/D1b* and cucumber *Csdw* (Wei et al., 2019). Fourteen dwarf mutants with short internodes and small leaf sizes were explored in this study. Previous studies reported that plant height-related genes are associated with leaf sizes such as *OsNAL1*, *OsNAL7*, and *OsDNL-4* (Bae et al., 2021). Similar to recent studies, six mutants of line 681, 942, 1372, 1833, 2194, and 3210 have short plant heights and small leaf sizes.

Fruit phenotype, including fruit shape, size, color, and surface texture possibly caused by long-term domestication, artificial selection, natural selection, or environmental factors, are crucial characteristics for different markets and consumers, and the variety of fruit gives rise to uncovering the molecular mechanism and genetic basis (Snouffer et al., 2020). For example, in bitter gourd, the fruit of wild type with tubercles is small, round, or spindle, while the cultivated species are larger, longer, and do not taper at both ends (Behera et al., 2010). In the mutant collection, we found small, slender, and bending fruits (Figures 6E–G). Moreover, we also found some mutants with different tubercles like flat, grain, and stripe warts on the skin (Figures 6B–D). Some studies have proved vital genes or QTLs that control fruit shapes, such as *CsFUL* and *CsTRM5* in cucumber, *CmOFP1a* in melon, and *ClSUN25-26-27a* in watermelon (Pan et al., 2020; Snouffer et al., 2020; Boualem et al., 2022). Furthermore, some mutants with a good fruit shape, that do not taper at both ends, and meet our breeding objectives were kept for future study. Therefore, further molecular analysis is needed and important to discover these genes controlling related fruit traits.

Conclusion

In this study, we confirmed the optimal EMS mutagenesis conditions for bitter gourd seeds of Y52 (EMS concentration 0.2% and mutagenic time 10 h). Furthermore, a mutant collection including 3223 mutants was constructed. We classified them according to the detection of phenotype characters into leaf shape mutation, leaf color mutation, dwarfing mutation, apical meristem mutation, fruit size, and shape mutation. These mutants not only bring new germplasms for breeding but can also accelerate gene functional study in bitter gourd.

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

RY, YY, and XW conceived and designed the project. RY and XW conducted the EMS mutation experiment, RY and YN collected most of the phenotype. KY, XH, ZL and ZQ participated in some mutant phenotype collected. RY and XW wrote the manuscript, and YY revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grants from the Hainan Province Science and Technology Special Fund (ZDYF2020063), The Major Science and Technology plan of Hainan Province (ZDKJ2021010), Hainan Yazhou Bay Seed Lab (B21HJ0305), and Hainan Province Science and Technology Special Fund (ZDYF2020059).

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.2022.1069750/full#supplementary-material

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30 (2), 174–178. doi: 10.1038/nbt.2095

Arisha, M. H., Shah, S. N., Gong, Z. H., Jing, H., Li, C., and Zhang, H. X. (2015). Ethyl methane sulfonate induced mutations in M2 generation and physiological variations in M1 generation of peppers (Capsicum annuum l.). *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00399

Bae, K. D., Um, T. Y., Yang, W. T., Park, T. H., Hong, S. Y., Kim, K. M., et al. (2021). Characterization of dwarf and narrow leaf (dnl-4) mutant in rice. *Plant Signal Behav.* 16 (2), 1849490. doi: 10.1080/15592324.2020.1849490

Behera, T. K., Behera, S., and Bharathi, L. K.. (2010). Bitter gourd: botany, horticulture, breeding. *Hortic. Rev.* 101–141. doi: 10.1002/9780470543672.ch2

Boualem, A., Berthet, S., Devani, R. S., Camps, C., Fleurier, S., Morin, H., et al. (2022). Ethylene plays a dual role in sex determination and fruit shape in cucurbits. *Curr. Biol.* 32 (11), 2390–2401.e2394. doi: 10.1016/j.cub.2022.04.031

Cui, J., Cheng, J., Nong, D., Peng, J., Hu, Y., He, W., et al. (2017). Genome-wide analysis of simple sequence repeats in bitter gourd (Momordica charantia). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01103

Cui, J., Luo, S., Niu, Y., Huang, R., Wen, Q., Su, J., et al. (2018). A RAD-based genetic map for anchoring scaffold sequences and identifying QTLs in bitter gourd (Momordica charantia). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00477

Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol. Plant* 15 (8), 1268–1284. doi: 10.1016/j.molp.2022.06.010

Dutta, S., Hazra, P., Saha, S., Acharya, B., Bhattacharjee, T., Kumar Maurya, P., et al. (2021). Applied mutagenesis could improve economically important traits in bitter gourd (Momordica charantia l.). *J. Genet.* 100, 34282734. doi: 10.1007/s12041-021-01283-5

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in arabidopsis. *Genetics* 164 (2), 731–740. doi: 10.1093/genetics/164.2.731

Henry, I. M., Nagalakshmi, U., and Lieberman, M. C. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* 26 (4), 1382–1397. doi: 10.1105/tpc.113.121590

Jander, G., Baerson, S. R., Hudak, J. A., Gonzalez, K. A., Gruys, K. J., and Last, R. L. (2003). Ethylmethanesulfonate saturation mutagenesis in arabidopsis to determine frequency of herbicide resistance. *Plant Physiol.* 131 (1), 139–146. doi: 10.1104/pp.102.010397

Kaur, G., Pathak, M., Singla, D., Chhabra, G., Chhuneja, P., and Kaur Sarao, N. (2021). Quantitative trait loci mapping for earliness, fruit, and seed related traits using high density genotyping-by-Sequencing-Based genetic map in bitter gourd (Momordica charantia l.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.799932

Lian, X., Liu, Y., Guo, H., Fan, Y., Wu, J., Guo, H., et al. (2020). Ethyl methanesulfonate mutant library construction in gossypium hirsutum l. for

allotetraploid functional genomics and germplasm innovation. *Plant J.* 103 (2), 858–868. doi: 10.1111/tpj.14755

Li, Z., Jiang, L., Ma, Y., Wei, Z., Hong, H., Liu, Z., et al. (2017). Development and utilization of a new chemically-induced soybean library with a high mutation density. *J Integr. Plant Biol.* 59 (1), 60–74. doi: 10.1111/jipb.12505

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38 (5), 582–585. doi: 10.1038/s41587-020-0455-x

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 10, 30272626. doi: 10.1038/nbt.4273

Lu, X., Liu, J., Ren, W., Yang, Q., Chai, Z., Chen, R., et al. (2018). Gene-indexed mutations in maize. *Mol. Plant* 11 (3), 496–504. doi: 10.1016/j.molp.2017.11.013

Ma, X., Sun, X., Li, C., Huan, R., Sun, C., Wang, Y., et al. (2017). Map-based cloning and characterization of the novel yellow-green leaf gene ys83 in rice (Oryza sativa). *Plant Physiol. Biochem.* 111, 1–9. doi: 10.1016/j.plaphy.2016.11.007

Pan, Y., Wang, Y., McGregor, C., Liu, S., Luan, F., Gao, M., et al. (2020). Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. *Theor. Appl. Genet.* 133 (1), 1–21. doi: 10.1007/s00122-019-03481-3

Qu, L. J., and Qin, G. (2014). Generation and identification of arabidopsis EMS mutants. *Methods Mol. Biol.* 1062, 225–239. doi: 10.1007/978-1-62703-580-4_12

Rao, P. G., Behera, T. K., Gaikwad, A. B., Munshi, A. D., and Boopalakrishnan, G. (2018). Mapping and QTL analysis of gynoecy and earliness in bitter gourd (Momordica charantia l.) using genotyping-by-Sequencing (GBS) technology. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01555

Rong, F., Chen, F., Huang, L., Zhang, J., Zhang, C., Hou, D., et al. (2019). A mutation in class III homeodomain-leucine zipper (HD-ZIP III) transcription factor results in curly leaf (cul) in cucumber (Cucumis sativus l.). *Theor. Appl. Genet.* 132 (1), 113–123. doi: 10.1007/s00122-018-3198-z

Snouffer, A., Kraus, C., and van der Knaap, E. (2020). The shape of things to come: ovate family proteins regulate plant organ shape. *Curr. Opin. Plant Biol.* 53, 98–105. doi: 10.1016/j.pbi.2019.10.005

Sun, X., Li, X., Lu, Y., Wang, S., Zhang, X., Zhang, K., et al. (2022). Construction of a high-density mutant population of Chinese cabbage facilitates the genetic dissection of agronomic traits. *Mol. Plant* 15 (5), 913–924. doi: 10.1016/j.molp.2022.02.006

Sun, Y., Zhang, H., Fan, M., He, Y., and Guo, P. (2020). A mutation in the intron splice acceptor site of a GA3ox gene confers dwarf architecture in watermelon (Citrullus lanatus l.). *Sci. Rep.* 10 (1), 14915. doi: 10.1038/s41598-020-71861-7

Wang, K., Shi, L., Liang, X., Zhao, P., Wang, W., Liu, J., et al. (2022). The gene TaWOX5 overcomes genotype dependency in wheat genetic transformation. *Nat. Plants* 8 (2), 110–117. doi: 10.1038/s41477-021-01085-8

Wei, C., Zhu, C., Yang, L., Zhao, W., Ma, R., 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. *Hortic. Res.* 6, 132. doi: 10.1038/s41438-019-0213-8

Yang, J., Weng, Y., Li, H., and Kong, Q. (2022). Epidermal patterning factor 2-like (McEPFL2): A putative candidate for the continuous ridge (cr) fruit skin locus in bitter gourd (*Momordica charantia* l.). *Genes (Basel)* 13 (7), 35885929. doi: 10.3390/genes13071148

Zhang, M., Song, M., Davoudi, M., Cheng, F., Yin, J., Zha, G., et al. (2022). The mutation of c-24 reductase, a key enzyme involved in brassinolide biosynthesis, confers a novel compact plant architecture phenotype to cucumber. *Theor. Appl. Genet.* 135 (8), 2711–2723. doi: 10.1007/s00122-022-04144-6

Zhang, X., Wang, Y., Zhu, X., Wang, X., Zhu, Z., Li, Y., et al. (2020). Curled flag leaf 2, encoding a cytochrome P450 protein, regulated by the transcription factor Roc5, influences flag leaf development in rice. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.616977

Zhong, J., Cheng, J., Cui, J., Hu, F., Dong, J., Liu, J., et al. (2022). MC03g0810, an important candidate gene controlling black seed coat color in bitter gourd (Momordica spp.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.875631

Zhu, H., Zhang, M., Sun, S., Yang, S., Li, J., Li, H., et al. (2019). A single nucleotide deletion in an ABC transporter gene leads to a dwarf phenotype in watermelon. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01399

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 December 2022
DOI 10.3389/fpls.2022.1074145

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Hao Li,
Northwest A&F University, China
Zhongyuan Hu,
Zhejiang University, China
Yongbing Zhang,
Hami Melon Research Centre, China

*CORRESPONDENCE
Nan He
henan@caas.cn
Wenge Liu
liuwenge@caas.cn

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

RECEIVED 19 October 2022
ACCEPTED 17 November 2022
PUBLISHED 06 December 2022

CITATION
Gong C, Lu X, Zhu H, Anees M, He N and Liu W (2022) Genome-wide association study provides genetic insights into natural variation in watermelon rind thickness and single fruit weight.
Front. Plant Sci. 13:1074145.
doi: 10.3389/fpls.2022.1074145

Genome-wide association study provides genetic insights into natural variation in watermelon rind thickness and single fruit weight

Chengsheng Gong[1,2], Xuqiang Lu[1], Hongju Zhu[1], Muhammad Anees[1], Nan He[1]* and Wenge Liu[1]*

[1]Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China, [2]Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Rind thickness and fruit weight are agronomic traits closely related to quality and yield, which have attracted much attention from consumers and breeders. However, the genetic mechanism of these two traits is still not well understood in natural populations. In this study, rind thickness and single fruit weight in 151 watermelon accessions were determined in 2019 and 2020, and genome-wide association analysis was performed by integrating phenotypic and genotype data. Abundant phenotypic variation was found in the test population, and the watermelon with thinner rind thickness tended to have smaller fruit weights. Five significant SNPs were closely associated with rind thickness on chromosome 2 by Genome-wide association study (GWAS), i.e., 32344170, 32321308, 32304738, 32328501, and 32311192. And there were 21 genes were annotated in the candidate interval, most notably, *Cla97C02G044160* belonged to the MADS family, and part of the genes in this family played an important role in regulating organ size, further analysis of gene structure, gene expression level, and phylogenetic tree showed that *Cla97C02G044160* was a candidate gene for regulating target traits. In conclusion, our study provides molecular insights into the natural variation of watermelon rind thickness and single fruit weight, meanwhile, providing data support for molecular marker-assisted breeding.

KEYWORDS

watermelon, GWAS, rind thickness, single fruit weight, transcriptome

Introduction

As one of the five most popular consumed fresh fruits, watermelon [*Citrullus lanatus* (Thunb.) Matsum. &Nakai var. lanatus] is widely cultivated in the world and brought considerable economic benefits (Guo et al., 2013; Gong et al., 2021b). With the increasing market demand for high-quality watermelons, research on important phenotypic traits such as fruit shape, flesh color, fruit sugar content, and metabolic traits has been deepened (Ren et al., 2017; Dou et al., 2018; Zhang et al., 2020; Gong et al., 2021a). Rind thickness and single fruit weight are two important agronomic traits that have attracted much attention in the market. Specifically, small-fruit watermelon with thinner rind tends to have better flavor and is an important component of high-quality watermelon in the market, while large-fruit watermelon with thicker rind thickness contributes to the increase of watermelon yield and transportation.

The research on the genetic basis of important agronomic traits is helpful to accelerate the modern breeding process (Ren et al., 2014). Forward genetics carries out genetic basis research based on phenotypic variation, and the relevant research is often carried out by genetic map, Bulked Segregant Analysis (BSA), and Genome-wide association study (GWAS). Advances have been made in gene mapping of single fruit weight and rind thickness through forwarding genetics, and key QTL has been obtained in cucumber, melon, watermelon, and other Cucurbitaceae crops (Pan et al., 2020). For instance, Zhong et al. (2017) obtained QTL loci linked to 12 important agronomic traits such as rind thickness based on the construction of a high-density genetic map of pumpkin, providing theoretical support for mining candidate genes and molecular marker breeding. Research on QTL linked to fruit thickness in wax gourds has also made progress (Liu et al., 2018). In watermelon, (Wi et al. 2012) obtained QTLs linked to fruit size, rind thickness, and other traits through different genetic populations; Ren et al. (2014) realized the main QTL loci for 12 agronomic traits such as fruit length and sugar content by integrating four genetic maps, including QTLs related to rind thickness and fruit weight, i.e., *RTH2-1*, *RTH5*, *RTH6*, *FWT2-1*, *FWT2-2*, *FWT3*, *FWT5-1*, and *FWT5-2*. In addition, forward genetics combined with transcriptome data is helpful for further mining and identifying key candidate genes that regulate target traits. Compared with genetic map and BSA, GWAS could be used to analyze the association between polymorphic nucleic acids and traits in a constructed population, and quickly identify target loci and possible regulatory genes associated with traits (Hamblin et al., 2011; Yano et al., 2016).

At present, important progress has been made in the research on the physiological and genetic basis of plant fruit weight and rind thickness, and some key candidate genes have been identified. The size of normal organs is usually affected by the coordination of cell proliferation and cell expansion (Horiguchi et al., 2006). Genes can regulate the weight of fruit, rind thickness, and the size of other traits by regulating the content of auxin, ethylene, cytokinin, and other plant hormones (Le et al., 2001; Wilmoth et al., 2005; Di Marzo et al., 2020). In addition, transcription factors (TFs) can also play an important role in regulating the size of plant tissues and organs, notably, the MADS family belongs to TFs and plays an important role in regulating the size of tissues and organs. For instance, in Arabidopsis, the TF of MADS gene *GORDITA* (GOA) has the function of regulating fruit growth, and mutation or reduced expression of this gene results in larger fruit (Prasad et al., 2010); *SEEDSTICK* regulating cytokinin levels and other genes to controls fruit weight of *Arabidopsis* (Di Marzo et al., 2020); In addition, rind thickness and fruit size have been extensively studied in tomato, and the genes such as *SlARF7* were considered to be key candidate genes related to regulating tomato rind thickness (De Jong et al., 2009). Furthermore, the construction of metabolic regulatory networks based on gene expression levels is helpful to elucidate the genetic basis of important agronomic traits (Umer et al., 2020). However, the current research on the key candidate genes regulating target traits is still far from enough.

Research on the genetic basis of single fruit weight and rind thickness in watermelon can help meet the practical needs of production, progress has been made in the study of rind thickness and fruit weight, while the understanding of the variation and genetic basis of these two traits in the natural population is still insufficient. In the present work, variation analysis and correlation analysis were carried out after the single fruit weight and rind thickness were measured of 151 watermelon accessions in 2019 and 2020. Combined with high-quality SNP data, GWAS of the target traits was performed and significant SNPs were obtained. Further, combine bioinformatics analysis and transcriptome data mining to identify key candidate genes that regulate target traits. In conclusion, our study provides molecular insights into the natural variation of rind thickness and single fruit weight and contributes to marker-assisted breeding and new cultivars selection.

Materials and methods

Plant cultivation and management

151 watermelon accessions belong to cultivated cultivars (*C. lanatus cultivar* and *C. lanatus landrace*), with bright flesh color to facilitate the identification of rind thickness. These accessions were randomly selected from previous studies with resequencing data (Guo et al., 2019; Gong et al., 2022), and were collected in the National Mid-term Genebank for Watermelon and Melon, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences. The plants were planted in Xinxiang (35.30°N, 113.88° E) in 2019 and Zhongmu (34.72°N, 113.98 E)

in 2020, respectively. The rind of 14 watermelon accessions with different thicknesses (grey background in Table S1) was sampled at the maturity period for qRT-PCR (Quantitative real-time polymerase chain reaction, qRT-PCR). In addition, two watermelon cultivars '97103' and '203Z' with different single fruit weights ('97103':3.2 kg, '203Z':2.0 kg) and rind thicknesses ('97103':1.1 cm, '203Z':0.5 cm) were mainly used for transcriptome determination and planted in Sanya, Hainan in the winter of 2021. And the plants were transplanted through seedling raising and according to the local climate conditions. Twin-vine pruning was used, and only one watermelon fruit per plant was retained. The main temperature and humidity in the greenhouse were as follows: during the daytime, the temperature was kept between 25°C to 35°C, and the humidity was 55%-70%, while at nighttime, the temperature was often above 15°C, and the humidity was maintained at 75% to 80%. Alternatively, irrigation, watering, and fertilization were managed according to the actual needs of watermelon growth and were managed in the same way. The cultivars were harvested according to their ripening characteristics, and then, the single fruit weight and rind thickness were measured.

Determination of phenotypic data and sampling

For the 151 watermelon accessions, three mature watermelons with good growth were randomly selected from each cultivar for the determination of target traits, and the phenotype data after calculating the average value were used for further analysis. The electronic scale was used to measure the weight of the fruit, and after the watermelon fruit was cut lengthwise, the thickness of the rind was measured with a ruler. Watermelon accessions '97103' and '203Z' were used to perform transcriptome sequencing, and 14 watermelon accessions were used to qRT-PCR, the rind of watermelon without exocarp was sampled at mature period, and the watermelon was cut vertically into 1 cm thick slices along the largest diameter, and cuboid pericarp of about 1 cm in length and width was taken from the four vertical directions and mixed as a sample to be tested. Three watermelons of the same growth were individually sampled as triplicate biological replicates.

GWAS and candidate gene identification

In the previous study, the linkage disequilibrium (LD) decay rates of the *C. lanatus* cultivar and *C. lanatus* landrace were relatively low (Guo et al, 2019), showing a high degree of domestication, and eliminating the influence of wild-type cultivars on genotype analysis due to differences in genetic background to a certain extent. The resequencing data of 126 watermelon accessions were obtained from (Guo et al. 2019) (accession number WM in Table S1), and the other 25 were obtained from Gong et al. (2022) (accession number R in Table S1). The method of gene library construction and resequencing was consistent with a previous study (Guo et al., 2019). TruSeq Nano DNA High Throughput Library Prep Kit was used to construct about 400 bp sequences. Sequencing was performed on the Illumina HiSeq X or HiSeq 2000 platform at Berry Genomics (Beijing, China). And the watermelon genome 97103 v2 (http://cucurbitgenomics.org/organism/21) was used as the reference genome for the related analysis, and QQ plots were plotted to predict the reliability of the model. The Manhattan plot and QQ plot were drawn by R software. The SNP data of mutation detection were filtered according to the secondary allele frequency (MAF: 0.05) and locus integrity (INT: 0.8), and the high-quality SNP was obtained for further analysis. For GWAS, a Factored Spectrally Transformed linear mixed model (FaST-LMM) method was used for relevant analysis, and QQ plots were plotted to predict the reliability of the model, and EMMA eXpedited(EMMAX) and Linear Model are used to verify the reliability of the location results. The Manhattan plot and QQ plot were drawn by R software. And the modified Bonferroni correction was used to determine two significance thresholds, i.e., -log10 (P) 0.1/Ne (Ne = effective SNP number) and -log10 (P) 0.01/Ne0.01/SNPs and 0.1/SNPs, respectively. Criteria for identifying significant SNPs: when the SNP exceeded the top threshold in the Manhattan plot, it was considered to be highly significantly associated with the target trait and SNP. When the lower threshold line was exceeded, the correlation between the two was considered significant. When located below the threshold line, SNPs that are close to the threshold line and show a continuous distribution in a small interval were also considered may be associated with the target trait.

According to the analysis of LD decay in natural populations in previous studies (Guo et al., 2019), we took the interval of 100 kb upstream and downstream of the significant SNP locus as the candidate interval. Furthermore, combined gene annotation information, previous research reports, and transcriptome data to screen possible candidate genes. Gene sequence information was obtained from the online website of the cucurbitaceous crop genome (http://www.cucurbitgenomics.org/).

Transcriptome sequencing and qRT-PCR

The rind samples of '97103' and '203Z' with three biological repeats were ground into powder in liquid nitrogen and then used for RNA extraction. RNA extraction was performed according to the instructions of the TRIzol Reagent (Life Technologies, California, USA). Qualified high-quality RNA was used for the next research, and NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530) and NEBNext Multiplex Oligos for Illumina (NEB, E7500) was used for cDNA library construction. The cDNA library was sequenced at an Illumina HiSeq™ sequencing platform. The original sequencing data were filtered

to obtain clean data, which were compared with the watermelon genome 97103 v2, and further quality control was performed to obtain high-quality data. FPKM values (number of exon kilobases per million segments) were used to estimate gene expression levels (Florea et al., 2013). When the fold change of FPKM value between the two cultivars exceeded 1.8, the gene was defined as a differential gene. According to previous research reports, the genes related to seed expansion and hormones in watermelons were counted, and the heat map was drawn based on the expression level of Origin software (V 8.0.724).

Fourteen watermelon cultivars were set with three biological replicates. The powdered watermelon rind samples were extracted with the RNA isolation kit (Huayueyang Biotechnologies, China) for RNA extraction, and then synthesis of cDNA with NovoScript plus all-in-one 1st stand cDNA synthesis supermix (Novoprotein, China). Mix the solution to be tested using the SYBR Green real-time PCR mix, and an instrument LightCycler480 RT-PCR system (Roche, Swiss) was used for measurement. *ClCAC* (Gene ID: *Cla97C09G174930*) was used as the reference gene. The original data obtained are calculated as the relative expression level by the $2-\Delta\Delta^{CT}$ method (Livak and Schmittgen, 2001).

Functional analysis of candidate genes

The construction of systematic evolution process: Amino acid sequence of the target gene was obtained from the watermelon reference genome 97103 v2. NCBI BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to obtain homologous genes of the target gene, database selection 'UniProtKB/Swiss-Prot (Swissprot)', and then download the full-length sequence of amino acids for further analysis. MEGE7 (7.0.26) was used for sequence alignment and phylogenetic tree construction, ClustalW was used for amino acid sequence alignment, and Construct/Test neighbor-joining Tree was used for phylogenetic tree construction. The conservative domain was predicted by NCBI Conserved Domains: https://www.ncbi.nlm.nih.gov/Structure/cdd/docs/cdd_search.html; the transmembrane structure analysis was predicted through an online website: TMHMM: http://www.cbs.dtu.dk/services/TMHMM/.

Results

Natural variation and correlation analysis of rind thickness and single fruit weight

In 2019 and 2020, the coefficient of variation of rind thickness in 151 watermelon accessions was 26.17% and 27.20%, respectively, while the coefficient of variation of single fruit weight was 28.01% and 23.02%, which suggests that these cultivated cultivars had abundant variations in these two agronomic traits. In the process of agricultural production, 0.6cm is an important reference to determine whether watermelon has a thicker rind, statistical analysis showed that there were 14 and 15 accessions with a rind thickness of less than 0.6 cm, 128 and 132 accessions between 0.6-1.6 cm, and 9 and 4 accessions of more than 1.6 cm in two years, respectively. For single fruit weight, there were 19 and 11 accessions with less than 2 kg, 111 and 122 accessions of 2-4 kg, meanwhile, 21 and 18 accessions with more than 4 kg, respectively (Table S1). At the same time, most of the kurtosis and skewness of rind thickness and single fruit weight was less than 1 in two years (Table 1) and combined with the statistical histogram (Figures 1A, B, Figures 1D, E), it was concluded that these two important agronomic traits were basically in line with the characteristics of normal distribution in the population. Correlation analysis showed that epidermal thickness had a significant positive relationship in both years (Figure 1C), as did single fruit weight (Figure 1F), implying that phenotypic differences may be regulated by both genes and the environment, and were the relatively stable quantitative traits.

To explore the relationship between rind thickness and single fruit weight, a correlation analysis between the two traits was performed. The results indicate that there was a positive correlation between rind thickness and single fruit weight, and the correlation values of the two traits were 0.51 and 0.47 in 2019 and 2020, respectively (Figure S1). Notably, when the rind thickness was less than 0.6, the average single fruit weight was 2.09 kg (SD=0.53) and 2.40 kg (SD=0.59), respectively; while when the rind thickness was greater than or equal to 0.6, the fruit weigh was 3.16 kg (SD=0.82) and 3.29 kg (SD=0.70). This difference indicates that there was a positive correlation between rind thickness and single fruit weight, especially for the watermelon with thinner rind thickness, which tends to have smaller fruits.

GWAS and significant SNP acquisition of target traits

Significant SNPs were obtained by GWAS for rind thickness and single fruit weight in 2019 and 2020 (Figure 2), and the QQ plot indicates the reliability of the GWAS analytical model (Figure S2). The analysis by the FAST-LMM model of rind

TABLE 1 Statistical analysis of rind thickness and single fruit weight of 151 watermelon accessions in 2019 and 2020.

Trait	Mean	SD	CV/%	Kurtosis	Skewness
2020 RTH	1.07	0.29	0.27	0.06	-0.54
2019 RTH	1.01	0.26	0.26	0.98	0.08
2020 FWT	3.21	0.74	0.23	1.15	0.13
2019 FWT	3.06	0.86	0.28	0.05	0.34

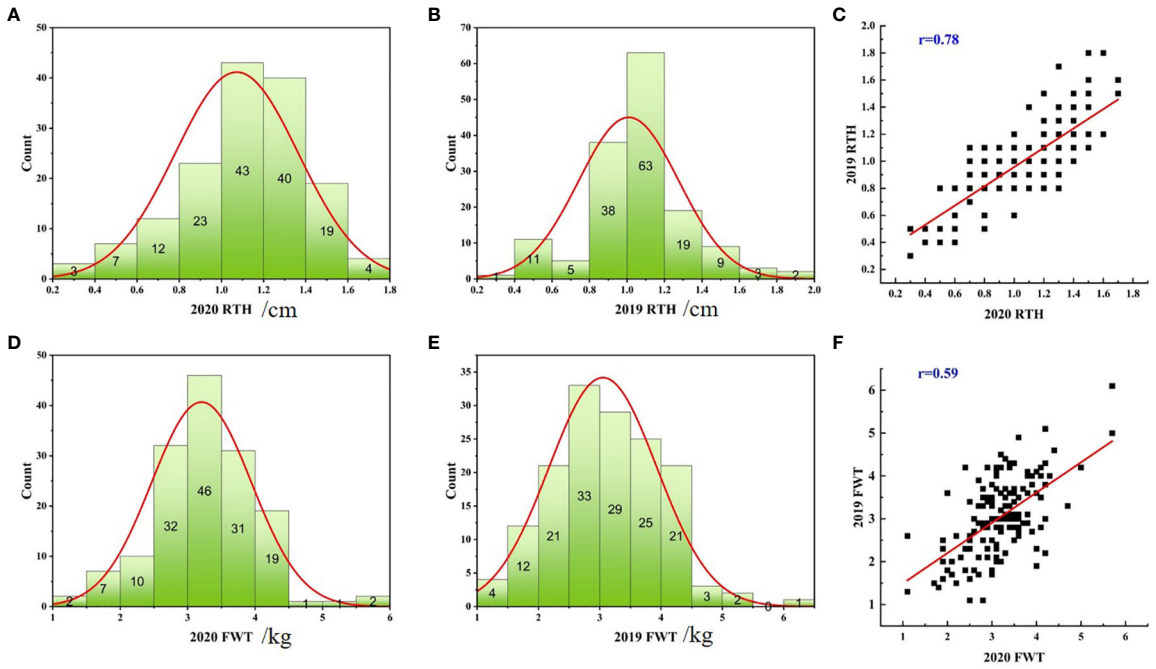


FIGURE 1
Distribution histograms and correlation graphs of rind thickness and single fruit weight of 151 watermelon accessions in 2019 and 2020. Statistical histogram of the rind thickness in 2020 **(A)**, the rind thickness in 2019 **(B)**, the single fruit weight in 2020 **(D)** and the single fruit weight in 2019 **(E)**, respectively. Correlation diagram of two-year data of rind thickness **(C)** and single fruit weight **(F)**, respectively.

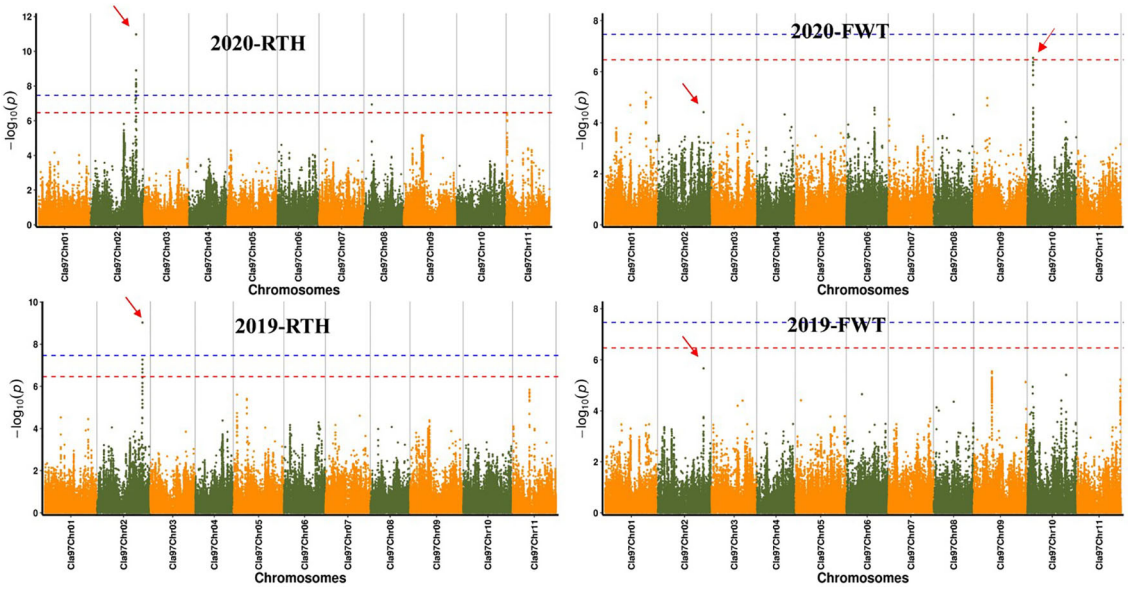


FIGURE 2
Manhattan plot of watermelon rind thickness and single fruit weight by GWAS in 2019 and 2020. The abscissa represents the chromosome, the ordinate represents -$\log_{10}$ (p-value), and the red and blue line segments parallel to the coordinate axis represent the two threshold lines, SNPs over red represent highly significant positive associations with traits, SNPs over blue are significantly associated with traits, respectively.

thickness found that there were 5 and 16 SNPs that exceeded the threshold line in two years (Table S2), and five same significant SNPs (S2: 32344170, S2: 32321308, S2: 32304738, S2: 32328501, S2: 32311192) on chromosome 2 were located on chromosome 2 in 2019 and 2020, and these loci were within the range of 50 kb. Besides, S8:5576295 was a significant SNP of rind thickness on chromosome 8 in 2020, and S10:4123404 was a significant SNP of single fruit weight on chromosome 10. Interestingly, S2: 32344170, which was significantly associated with rind thickness, was also the most significant SNP on chromosome 2 for single fruit weight. It is worth noting that GWAS analysis by EMMAX and LM has relatively similar positioning results with FAST-LMM, which further explains the reliability of our positioning results (Figure S3).

The acquisition of candidate interval and candidate gene

Notably, through the above analysis, we found that significant SNP loci were within a candidate interval of about 50 kb. Interestingly, S2: 32304738 and S2: 32311192 were located on the promoter and intron of *Cla97C02G044120*, respectively, and S2: 32321308 and S2: 32328501 were located on the intron and exon of *Cla97C02G044130*, respectively, and S2: 32344170 was the most significant SNP. To obtain allele information of significant SNP based on resequencing data, it is worth noting that when these alleles were in mutant alleles, they have thinner rind thickness and smaller fruit, and this difference was more obvious at S2: 32344170 (Figure 3). Moreover, it was found that when the rind thickness was less than 0.6 cm, all the alleles correspond to the mutant alleles (Table S1). These results further illustrate the close correlation between significant SNP with target traits on chromosome 2, and S2: 32344170 could be used as potential molecular markers for marker-assisted breeding of target traits. Furthermore, the correlation analysis of regions and genes near the most significant locus S2: 32344170 was analyzed, and the 100 kb upstream and downstream intervals of this site were defined as candidate intervals for further analysis. LD block analysis found that SNP in the candidate interval had a relatively good correlation coefficient (Figure S4), indicating that there may be a linkage imbalance in the candidate interval. Interestingly, marker NW0249226 in the previous QTL *Qrth2-1* for rind thickness was only about 50 kb away from S2: 32344170.

A total of 21 genes were obtained in the candidate interval (Table 2). GO annotation analysis found that in the molecular_function term, more genes were enriched in GTP binding, molecular_function, protein binding, ATP binding, protein kinase activity, and transferase activity; in the cellular_component term, an integral component of membrane, membrane, nucleus, and plasma membrane had more genes were annotated; in biological_process term, more genes were enriched in protein phosphorylation, MAPK cascade, regulation of transcription, DNA-templated phosphorylation had annotated more genes. It is worth noting that protein phosphorylation, DNA-templated phosphorylation, etc. were often considered to be involved in auxin biosynthesis.

Gene expression analysis of *Cla97C02G044160*

Combined with transcriptome data in '97103' and '203Z', the genes in the candidate interval were analyzed. It was found that

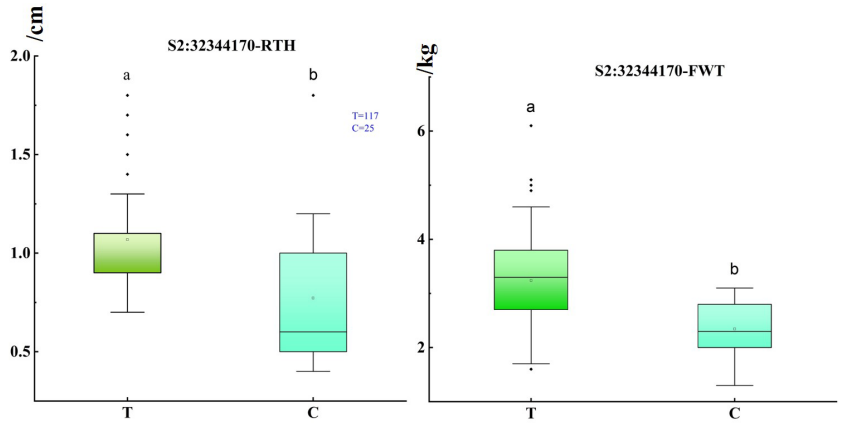


FIGURE 3
Box plot of single fruit weight and rind thickness under different bases of S2: 32344170. The abscissa represents the alleles with a mutation at the locus, the reference base near the origin, the other one was the key mutation locus, and the ordinate represents the rind thickness and fruit weight, respectively. Significance level: 0.05, a, and b represented significant differences in rind thickness under different alleles. The hollow origin in the boxplot represents the mean value.

TABLE 2 Candidate gene information of rind thickness and single fruit weight on chromosome 2.

Gene	Location	Gene annotation
Cla97C02G044060	Cla97Chr02:32246890.32248741 (+)	transcription elongation factor B polypeptide 3-like
Cla97C02G044070	Cla97Chr02:32252262.32256301(-)	Citrate synthase
Cla97C02G044080	Cla97Chr02:32264788.32269245(-)	Ubiquinone biosynthesis O-methyltransferase, mitochondrial
Cla97C02G044090	Cla97Chr02:32278703.32280043(-)	Leucine-rich repeat receptor-like protein kinase family protein
Cla97C02G044100	Cla97Chr02:32283295.32286529(-)	Serine/threonine protein phosphatase 7 long form
Cla97C02G044110	Cla97Chr02:32290370.32300164(-)	Serine/threonine-protein phosphatase
Cla97C02G044120	Cla97Chr02:32305922.32314246 (+)	elongation factor 2
Cla97C02G044130	Cla97Chr02:32320831.32328837 (+)	Protein kinase, putative
Cla97C02G044140	Cla97Chr02:32332095.32335291 (+)	elongation factor 2
Cla97C02G044150	Cla97Chr02:32337553.32338449(-)	HVA22-like protein
Cla97C02G044160	Cla97Chr02:32349097.32356436 (+)	MADS box transcription factor AGAMOUS
Cla97C02G044170	Cla97Chr02:32360188.32365056(-)	Hexosyltransferase
Cla97C02G044180	Cla97Chr02:32382369.32383409 (+)	Secretion-regulating guanine nucleotide exchange factor
Cla97C02G044190	Cla97Chr02:32384480.32385393(-)	protodermal factor 1-like
Cla97C02G044200	Cla97Chr02:32387278.32388270(-)	peroxisomal and mitochondrial division factor 2
Cla97C02G044210	Cla97Chr02:32390210.32394734(-)	Mitogen-activated protein kinase
Cla97C02G044220	Cla97Chr02:32411247.32412302(-)	glucan endo-1,3-beta-glucosidase 12-like
Cla97C02G044230	Cla97Chr02:32427610.32430022 (+)	WRKY transcription factor 100
Cla97C02G044240	Cla97Chr02:32433191.32433619(-)	Unknown protein
Cla97C02G044250	Cla97Chr02:32434065.32437213(-)	Receptor-like protein kinase
Cla97C02G044260	Cla97Chr02:32437502.32438686 (-)	Unknown protein

among the detected genes, only the gene expression level of *Cla97C02G044160* was more than 1.8 times between the two cultivars (Figure 4A). The analysis of this gene showed that *Cla97C02G044160* belongs to MADS family, and the typical conserved domains of MADS_MEF2 and K-Box, and transmembrane domain analysis found that K-Box region has a transmembrane region (Figure S5). Further combined with phylogenetic tree analysis, it was found that the gene had high homology with the previously reported genes *Osmads34*, *Osmads3*, and *Osmads13* regulating rice organ size (Figure S6) (Li et al., 2011; Zhang et al., 2016).

To fully understand the expression of candidate gene *Cla97C02G044160* in the natural population, RNA was extracted from 14 germplasm with different watermelon rind

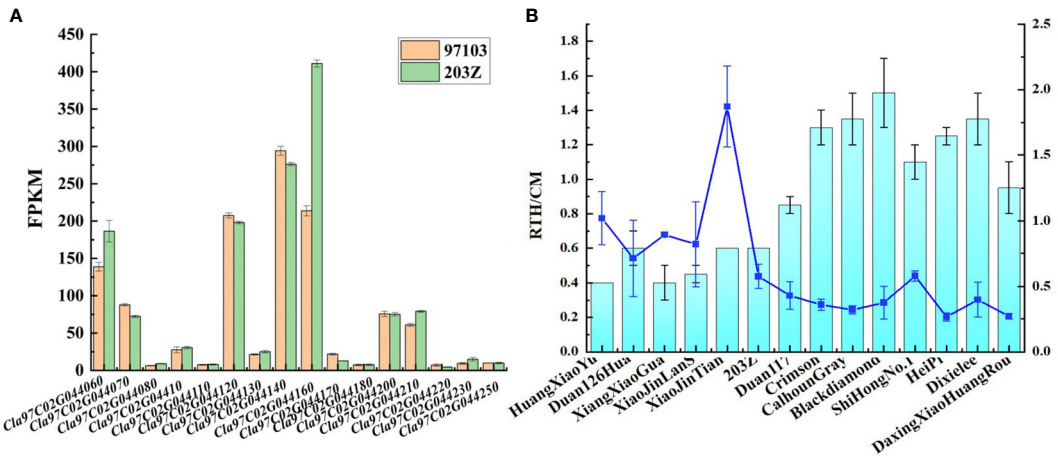


FIGURE 4
Expression levels of candidate genes for rind thickness and single fruit weight. **(A)** Statistical diagram of gene expression levels among '97103' and '203Z' within the candidate interval. **(B)** Rind thickness and relative gene expression level of *Cla97C02G044160* in 14 watermelon accessions.

thickness and fruit weight and further used for the qRT-PCR assay. The results showed that the gene expression level was significantly higher when the rind thickness was less than 0.6 cm than when the rind thickness was more than 0.6 cm. Specifically, The gene expression level of *Cla97C02G044160* in 'HuangXiaoYu', 'Duan126Hua', 'XiangXiaoGua', 'XiaoJinLanS', 'XiaoJinTian', '203Z' was significantly higher than that of 'Duan117', 'Crimson', 'CalhounGray', 'Blackdiamond', 'ShiHongNo.1', 'HeiPi', 'Dixielee' and 'DaxingXiaoHuangRou' (Figure 4B). It should be noted that this trend of gene expression was significantly opposite to that of rind thickness, that is, the gene expression level of *Cla97C02G044160* in rind was higher in watermelon with thinner rind. These results suggest that *Cla97C02G044160* maybe was a key candidate gene for regulating watermelon rind thickness and single fruit weight.

Transcriptome analysis of cultivars with different phenotypes

The genes whose expression was more than 1.8 times different between '97103' and '203Z' were counted. A total of 12 918 genes were screened for genes with an average content greater than 5. Furthermore, a total of 2117 genes were obtained by screening differentially expressed genes between the two cultivars. These 2117 genes were mainly divided into two categories, one has a higher gene expression in '97103', there are 1241 genes, and the second group had higher gene expression in '203Z', with 876 genes.

Among these genes, the statistics of genes related to organ expansion revealed that the expression levels of xyloglucan endotransglucosylase/hydrolase, cellulose synthase, pectinesterase, and pectinesterase inhibitor were higher in the watermelon '97103' with thicker rind thickness and heavier fruit; Most of the key genes auxin efflux carrier component, auxin transporter-like protein, ABC transporter B family member related to auxin synthesis had relatively higher gene expression levels in '97103'; as well as ethylene synthesis-related genes S-adenosylmethionine, 1-aminocyclopropane-1-carboxylate synthase family protein, 1-aminocyclopropane-1-carboxylate oxidase 1, most of them have higher gene expression levels in '97103'; while the cytokinin-related genes cytokinin riboside 5'-monophosphate phosphoribohydrolase, adenylate isopentenyltransferase have higher content in '203Z' (Figure 5). Combined with the GO annotation information in the candidate interval based on GWAS, it is speculated that the genes related to the regulation of auxin synthesis were likely to regulate target trait. Combined with the analysis of the expression level of candidate gene *Cla97C02G044160*, these genes with the same or opposite gene expression levels may have a potential regulatory relationship with *Cla97C02G044160*, or may have synergistic effects to make watermelon show different single fruit weight or rind thickness.

Discussion

Single fruit weight and rind thickness of watermelons are important agronomic traits closely related to domestication and yield. Bahari et al. (2012) screened watermelon cross combinations suitable for commercial cultivation by calculating the general combining ability (GCA) and special combining ability (SCA) of agronomic traits such as rind thickness and fruit weight, to screen out the hybrid combination more in line with the market demand. In the process of domestication, organ expansion is one of the important characteristics, Guo et al. (2019) found the average fruit weight of *C. Colocynthis*, *C. amarus*, *C. mucosospermus*, and *C. lanatus landraces* and cultivars was 0.2, 1.7, 3.7 and 3.4 kg, respectively. However, the artificial selection makes modern cultivated cultivars have a higher degree of domestication, correlation analysis of 151 modern cultivated watermelons showed that the watermelons with thinner rinds tended to have smaller fruits, thus establishing a link between phenotypes. In addition, the abundant variety of watermelon natural populations provides more potential parents for future breeding work.

For gene mapping of quantitative traits, the genetic map is one of the methods to obtain major QTLs for target traits. For instance, Ren et al. (2014) integrated four previously reported genetic maps into a new map of watermelon, and 58 QTLs for 12 traits were mapped into the new map. And interestingly, co-localization of rind thickness (*RTH2-1*) and fruit weight (*FWT2-2*) was found on chromosome 2 and chromosome 5. Moreover, *FWT2-2*, which controls the weight of watermelon fruit, was verified by selective clearance analysis of natural population accessions (Guo et al., 2019). GWAS can quickly obtain molecular markers and candidate genes that are significantly related to traits by detecting genetic variation polymorphisms of multiple individuals in the whole genome and analyzing their association with traits and made important progress in the study of important agronomic traits of watermelon (Dou et al., 2018; Ren et al., 2021). In the present study, 5 significant SNP loci on chromosome 2 were obtained in 2019 and 2020. Interestingly, the highest SNP locus S2: 32344170 was only about 50 kb away from QTL *FWT2-2* and *RTH2-1* (*RTH2-1* was included in QTL *FWT2-2)* (Ren et al., 2014), which regulated watermelon single fruit weight and rind thickness, indicating the reliability of the significant SNPs we obtained. The discovery of molecular markers can be applied to the molecular marker-assisted breeding of crops (Fang et al., 2021; Zhuo et al., 2021), in this study, we found that significant SNPs had significant differences in rind thickness under different bases. For example, under the

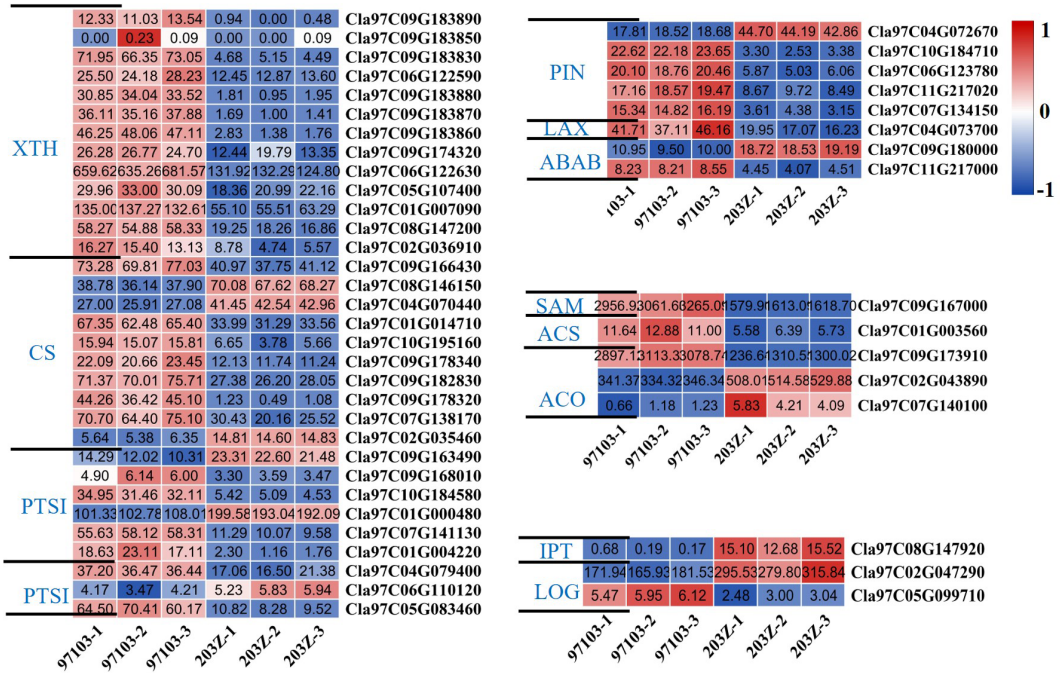


FIGURE 5
Heat map of differentially expressed genes related to seed expansion or hormone synthesis (auxin, ethylene, cytokinin). The darker the red color in the color scale, the higher the relative expression of the gene in the sample; the darker the blue color, the lower the expression of the gene in the sample. The value in the box represents the FPKM value in the sample. XTH, Xyloglucan endotransglucosylase/hydrolase; CS, Cellulose synthase; PTS, Pectinesterase; PTSI, Pectinesterase inhibitor; PIN, Auxin efflux carrier component; LAX, Auxin transporter-like protein; ABAB, ABC transporter B family member; SAM, S-adenosylmethionine; ACS, 1-aminocyclopropane-1-carboxylate synthase family protein; ACO, 1-aminocyclopropane-1-carboxylate oxidase 1; LOG, Cytokinin riboside 5′-monophosphate phosphoribohydrolase; IPT, Adenylate isopentenyltransferase.

highest significant SNP, when the rind thickness of cultivated cultivars was less than 0.6 cm, the corresponding base was all C. In summary, these significant SNPs can be used as potential molecular markers in watermelon breeding.

The differences in phenotypic traits are often regulated by genes. Combined with transcriptome data, we obtained the possible candidate gene *Cla97C02G044160* (a MADS family gene) in the candidate interval. In previous reports, genes in the MADS family can negatively regulate the size of tissues and organs by regulating hormones, etc. For instance, *SEEDSTICK* regulates cytokinin levels and other genes to control the fruit size of *Arabidopsis* (Di Marzo et al., 2020); RNA interference of SLMBP21 lead to longer sepals and improved fruit set in tomato, and maybe a significant correlation with auxin and ethylene (Li et al., 2017). In cucumber fruit, *CsFUL1A* inhibits the expression of auxin transporters *PIN-FORMED1* (PIN1) and *PIN7* and then resulting in decreases in auxin (Zhao et al., 2019). Zheng et al. (2020) found that two MADS-box transcription factors, *VCM1* and *VCM2*, play an important role in tree thickening. The discovery that the MADS family gene *Cla97C02G044160* could negatively regulate the size of tissues and organs was also consistent with this study, that is, the expression level of this gene in watermelon '203Z' with lighter fruit weight and the thinner rind was significantly higher than '97103'. In addition, combined with transcriptome data, it was further found that key genes related to organ expansion and auxin and ethylene had a higher accumulation amount in '97103', and our study further improved the transcriptional regulation basis of watermelon rind thickness and single fruit weight.

Conclusions

In the current study, we identified natural variations and correlations between rind thickness and single fruit weight in 151 watermelon accessions, and the significant SNP loci obtained by GWAS may serve as potential molecular markers to participate in breeding. Further combined with transcriptome data and bioinformatics analysis, it was found that *Cla97C02G044160* may be involved in the phenotypic changes of the target trait. Our research provides rich data resources for the selection of new cultivars and molecular marker-assisted breeding.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession numbers PRJNA641178 and PRJNA905076, all data is publicly available and can be used with the permission of the authors.

Author contributions

WL, NH, and CG designed this experiment. WL, XL, NH, and HZ were mainly responsible for the collection of watermelon accessions and plants management. WL and CG were mainly responsible for the statistics of phenotypic data. CG was mainly responsible for the data collation and analysis, and completed the writing of the original manuscript. WL, NH, and MA were mainly responsible for the manuscript verification. All the authors agreed to the submission of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2021-ZFRI), the China Agriculture Research System of MOF and MARA (CARS-25-03), National Key R&D Program of China (2018YFD0100704), the National Nature Science Foundation of China (31672178 and 31471893) and Natural Science Foundation of Henan Province (212300410312), and the Funding of Joint Research on Agricultural Variety Improvement of Henan Province (Grant No. 2022010503).

Acknowledgments

We sincerely thank the editors and reviewers for their contributions to the manuscript review. We also thank Henan Key Laboratory of Fruit and Cucurbit Biology for providing transcriptome detection and analysis in our experiment.

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.2022.1074145/full#supplementary-material

References

Bahari, M., Rafii, M. Y., Saleh, G. B., and Latif, M. A. (2012). Combining ability analysis in complete diallel cross of watermelon (*Citrullus lanatus* (Thunb.) matsum. *Nakai) ScientificWorldJournal* 2012, 543158–543158. doi: 10.1100/2012/543158

De Jong, M., Wolters-arts, M., Feron, R., Martiani, C., and Vriezen, W. H. (2009). The solanum lycopersicum auxin response factor 7 (SlARF7) regulates auxin signaling during tomato fruit set and development. *Plant J.* 57 (1), 160–170. doi: 10.1111/j.1365-313X.2008.03671.x

Di Marzo, M., Herrera-Ubaldo, H., Caporali, E., Novák, O., Strnad, M., Balanzà, V., et al. (2020). *SEEDSTICK* controls arabidopsis fruit size by regulating cytokinin levels and FRUITFULL. *Cell Rep.* 30, 2846–2857.e2843. doi: 10.1016/j.celrep.2020.01.101

Dou, J., Zhao, S., Lu, X., He, N., Zhang, L., Ali, A., et al. (2018). Genetic mapping reveals a candidate gene (*ClFS1*) for fruit shape in watermelon (Citrullus lanatus l.). *Theor. Appl. Genet.* 131, 947–958. doi: 10.1007/s00122-018-3050-5

Fang, K., Xia, Z., Li, H., Jiang, X., Qin, D., Wang, Q., et al. (2021). Genome-wide association analysis identified molecular markers associated with important tea flavor-related metabolites. *Hortic. Res.* 8, 42. doi: 10.1038/s41438-021-00477-3

Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res* 2, 188–188. doi: 10.12688/f1000research

Gong, C., Diao, W., Zhu, H., Umer, M. J., Zhao, S., He, N., et al. (2021a). Metabolome and transcriptome integration reveals insights into flavor formation of 'Crimson' watermelon flesh during fruit development. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.629361

Gong, C., Zhao, S., Yang, D., Lu, X., Anees, M., He, N., et al. (2022). Genome-wide association analysis provides molecular insights into natural variation in watermelon seed size. *Hortic. Res.* 9, uhab074. doi: 10.1093/hr/uhab074

Gong, C., Zhu, H., Lu, X., Yang, D., Zhao, S., Umer, M. J., et al. (2021b). An integrated transcriptome and metabolome approach reveals the accumulation of taste-related metabolites and gene regulatory networks during watermelon fruit development. *Planta* 254, 35. doi: 10.1007/s00425-021-03680-7

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470

Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4

Hamblin, M. T., Buckler, E. S., and Jannink, J.-L. (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27, 98–106. doi: 10.1016/j.tig.2010.12.003

Horiguchi, G., Ferjani, A., Fujikura, U., and Tsukaya, H. (2006). Coordination of cell proliferation and cell expansion in the control of leaf size in *Arabidopsis thaliana*. *J. Plant Res.* 119, 37–42. doi: 10.1007/s10265-005-0232-4

Le, J., Vandenbussche, F., van der Straeten, D., and Verbelen, J. P. (2001). In the early response of *Arabidopsis* roots to ethylene, cell elongation is up- and down-regulated and uncoupled from differentiation. *Plant Physiol.* 125, 519–522. doi: 10.1104/pp.125.2.519

Li, N., Huang, B., Tang, N., Jian, W., Zou, J., Chen, J., et al. (2017). The MADS-box gene *SlMBP21* regulates sepal size mediated by ethylene and auxin in tomato. *Plant Cell Physiol.* 58, 2241–2256. doi: 10.1093/pcp/pcx158

Li, H., Liang, W., Yin, C., Zhu, L., and Zhang, D. (2011). Genetic interaction of *OsMADS3*, DROOPING LEAF, and OsMADS13 in specifying rice floral organ identities and meristem determinacy. *Plant Physiol.* 56 (1), 263–274. doi: 10.1104/pp.111.172080

Liu, W., Jiang, B., Peng, Q., He, X., Lin, Y., Wang, M., et al. (2018). Genetic analysis and QTL mapping of fruit-related traits in wax gourd (*Benincasa hispida*). *Euphytica* 214 (8), 1–8. doi: 10.1007/s10681-018-2166-7

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

Pan, Y., Wang, Y., McGregor, C., Liu, S., Luan, F., Gao, M., et al. (2020). Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. *Theor. Appl. Genet.* 133, 1–21. doi: 10.1007/s00122-019-03481-3

Prasad, K., Zhang, X., Tobón, E., and Ambrose, B. A. (2010). The *Arabidopsis* b-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant J.* 62, 203–214. doi: 10.1111/j.1365-313X.2010.04139.x

Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., et al. (2017). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* 176, 836–850. doi: 10.1104/pp.17.01290

Ren, Y., Li, M., Guo, S., Sun, H., Zhao, J., Zhang, J., et al. (2021). Evolutionary gain of oligosaccharide hydrolysis and sugar transport enhanced carbohydrate partitioning in sweet watermelon fruits. *Plant Cell.* 33, 1554–1573. doi: 10.1093/plcell/koab055

Ren, Y., McGregor, C., Zhang, Y., Gong, G., Zhang, H., Guo, S., et al. (2014). An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol.* 14, 33. doi: 10.1186/1471-2229-14-33

Umer, M. J., Bin Safdar, L., Gebremeskel, H., Zhao, S., Yuan, P., Zhu, H., et al. (2020). Identification of key gene networks controlling organic acid and sugar metabolism during watermelon fruit development by integrating metabolic phenotypes and gene expression profiles. *Hortic. Res.* 7, 193–193. doi: 10.1038/s41438-020-00416-8

Wilmoth, J. C., Wang, S., Tiwari, S. B., Joshi, A. D., Hagen, G., Guilfoyle, T. J., et al. (2005). *NPH4/ARF7* and *ARF19* promote leaf expansion and auxin-induced lateral root formation. *Plant J.* 43, 118–130. doi: 10.1111/j.1365-313X.2005.02432.x

Wi Sandlin, K., Prothro, J., Heesacker, A., Khalilian, N., Okashah, R., Xiang, W., et al. (2012). Comparative mapping in watermelon [*Citrullus lanatus* (Thunb.) matsum. *Nakai] Theor. Appl. Genet.* 125, 1603–1618. doi: 10.1007/s00122-012-1938-z

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.-C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi: 10.1038/ng.3596

Zhang, J., Sun, H., Guo, S., Ren, Y., and Xu, Y. (2020). Decreased protein abundance of lycopene β-cyclase contributes to red flesh in domesticated watermelon. *Plant Physiol.* 183, 1171–1183. doi: 10.1104/pp.19.01409

Zhang, Y., Yu, H., Liu, J., Wang, W., Sun, J., Gao, Q., et al. (2016). Loss of function of OsMADS34 leads to large sterile lemma and low grain yield in rice (*Oryza sativa* l.). *Mol. Breed.* 36 (11), 147. doi: 10.1007/s11032-016-0578-4

Zhao, J., Jiang, L., Che, G., Pan, Y., Li, Y., Hou, Y., et al. (2019). A functional allele of *CsFUL1* regulates fruit length through repressing *CsSUP* and inhibiting auxin transport in cucumber. *Plant Cell.* 31, 1289–1307. doi: 10.1105/tpc.18.00905

Zheng, S., He, J., Lin, Z., Zhu, Y., Sun, J., and Li, L. (2020). Two MADS-box genes regulate vascular cambium activity and secondary growth by modulating auxin homeostasis in populus. *Plant Commun.* 23, 100134. doi: 10.1016/j.xplc.2020.100134

Zhong, Y. J., Zhou, Y. Y., Li, J. X., Yu, T., Wu, T. Q., Luo, J. N., et al. (2017). A high-density linkage map and QTL mapping of fruit-related traits in pumpkin (*Cucurbita moschata* duch.). *Sci. Rep.* 7 (1), 12785. doi: 10.1038/s41598-017-13216-3

Zhuo, X., Zheng, T., Li, S., Zhang, Z., Zhang, M., Zhang, Y., et al. (2021). Identification of the PmWEEP locus controlling weeping traits in prunus mume through an integrated genome-wide association study and quantitative trait locus mapping. *Hortic. Res.* 8, 131–131. doi: 10.1038/s41438-021-00573-4

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fpls.2022.1106123

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Peng Gao,
Northeast Agricultural University, China
Changlin Wang,
Insititute of Vegetables and Flowers (CAAS), China
Muhammad Azher Nawaz,
University of Sargodha, Pakistan

*CORRESPONDENCE
Dasen Xie
xiedasen@126.com

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

RECEIVED 23 November 2022
ACCEPTED 08 December 2022
PUBLISHED 22 December 2022

CITATION
Cai J, Yang S, Liu W, Yan J, Jiang B and Xie D (2022) A transcriptome analysis of *Benincasa hispida* revealed the pathways and genes involved in response to *Phytophthora melonis* infection.
Front. Plant Sci. 13:1106123.
doi: 10.3389/fpls.2022.1106123

A transcriptome analysis of *Benincasa hispida* revealed the pathways and genes involved in response to *Phytophthora melonis* infection

Jinsen Cai[1], Songguang Yang[1], Wenrui Liu[1], Jinqiang Yan[1], Biao Jiang[1] and Dasen Xie[1,2]*

[1]Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [2]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China

Wilt disease caused by *Phytophthora melonis* infection is one of the most serious threats to *Benincasa hispida* production. However, the mechanism of the response of *B. hispida* to a *P. melonis* infection remains largely unknown. In the present study, two *B. hispida* cultivars with different degrees of resistance to *P. melonis* were identified: B488 (a moderately resistant cultivar) and B214 (a moderately susceptible cultivar). RNA-seq was performed on *P. melonis*-infected B488 and B214 12 hours post infection (hpi). Compared with the control, 680 and 988 DEGs were respectively detected in B488 and B214. A KEGG pathway analysis combined with a cluster analysis revealed that phenylpropanoid biosynthesis, plant-pathogen interaction, the MAPK signaling pathway-plant, and plant hormone signal transduction were the most relevant pathways during the response of both B488 and B214 to *P. melonis* infection, as well as the differentially expressed genes in the two cultivars. In addition, a cluster analysis of transcription factor genes in DEGs identified four genes upregulated in B488 but not in B214 at 6 hpi and 12 hpi, which was confirmed by qRT-PCR. These were candidate genes for elucidating the mechanism of the *B. hispida* response to *P. melonis* infection and laying the foundation for the improvement of *B. hispida*.

KEYWORDS

RNA-Seq, *Benincasa hispida*, *Phytophthora melonis*, wilt disease, resistanceassociated gene

Introduction

In nature, interactions with microorganisms are unavoidable during the life-cycle of a plant. A large number of these microbes are plant pathogens; for example, pathogens in the genus *Phytophthora* (Chialva et al., 2022). *Phytophthora* belongs to the oomycete family, and some species, including *P. infestans*, *P. ramorum*, *P. sojae*, *P. nicotianae*, *P. capsica*, and *P. cinnamomi* are well-known pathogens that cause disease and significant losses in important agricultural and forestry crops worldwide. (Kamoun et al., 2015; Mora-Sala et al., 2022). Generally, *Phytophthora* species produce zoospores, which can reach speeds of up 250 mm/s when in swimming water (Appiah et al., 2005) before settling on the cell surface, germinating, and penetrating into the host cells. Under optimal conditions (i.e., 25-30°C and high relative humidity), a successful invasion is followed by growth and colonization of the host tissues, which eventually results in tissue collapse and sporulation, and this can turn into a serious epidemic within a few days or weeks (Lamour et al., 2012).

Pathogenic microbes destroy tissue and deprive host plants of nutrients, resulting in stunted growth and even death. Plants have developed defenses to survive pathogen attacks through long-term co-evolution with the pathogens (Zhou and Zhang, 2020; Jung et al., 2021). A plant has the capacity to recognize a diverse range of pathogen/microbe-associated molecular patterns (P/MAMP) *via* surface pattern-recognition receptors (PRRs), resulting in pattern-triggered immunity (PTI). Generally, PRRs are made up of receptor-like proteins and receptor kinases, which are found on the cell surface. Plants have also evolved resistance genes that encode nucleotide-binding leucine-rich repeat receptors, which recognize specific pathogen effectors, resulting in effector-triggered immunity (ETI) (Kourelis and van der Hoorn, 2018; Van de Weyer et al., 2019). The activation of immune receptors triggers downstream signaling, such as calcium influx, mitogen-activated protein kinase (MAPK) cascades, a burst of reactive oxygen species, accumulation of defense hormones, or the expression of immune marker genes (Tang et al., 2017; Tian et al., 2019; Huang et al., 2020). These effects can lead to defense execution processes such as callose deposition and the hypersensitive response (Jones and Dangl, 2006).

The activation of MAPK cascades is a major early signaling event in the plant defense response. There are two MAPK cascades; one is composed of MAP kinase kinase kinase (MEKK1), MAP kinase kinases (MKK1 and MKK2), and MAP kinase (MPK4) (Suarez-Rodriguez et al., 2007; Gao et al., 2008). A second cascade is composed of two MAPKKKs (MAPKKK3 and MAPKKK5), two MKKs (MKK4 and MKK5), and two MAPKs (MPK3 and MPK6) (Asai et al., 2002; Su et al., 2017). Additionally, positive feedback mechanisms are also involved, such as MAPKKK5, which is phosphorylated by MPK6 and further enhances the activation of MPK3/6, resulting in increased disease resistance (Bi et al., 2018). The receptors in the two layers (i.e., PTI and ETI) of the plant immune system detect MAMPs/effectors, which activate the MAPK cascades and trigger downstream signaling events such as plant hormone accumulation.

Plant hormones play a central role in plant immunity, and each different hormone regulates its own core pathway in the immune network. For instance, NPR1 and NPR3/NPR4, two classes of SA (salicylic acid) pathway receptors, interact with transcription factors (TGAs) to regulate transcriptional activation and repression (Zhang and Li, 2019). SA binds to receptors and induces defense gene expression by promoting NPR1 transcriptional activator activity and inhibiting NPR3/NPR4 transcriptional repression activity (Ding et al., 2018; Zhou and Zhang, 2020). The SA pathway is thought to be primarily directed at biotrophic pathogens (Zhang and Li, 2019). Additionally, JA (jasmonic acid) regulates another well-studied defense pathway (Wasternack and Song, 2017). The JA pathway can be subdivided into two branches: the ERF branch and the MYC branch. The ERF branch is co-regulated with ET (ethylene) and is associated with necrotrophic pathogens, while the MYC branch is co-regulated with ABA(abscisic acid) and provides general protection against chewing insects (Aerts et al., 2021).

In addition to signal molecules, secondary metabolites also play an important role in the plant defense response. The defensive functions of compounds produced by the phenylpropanoid pathways have been extensively researched. For example, lignification is caused by the deposition of lignin in the cell wall and serves as the first line of defense against pathogens (Bechinger et al., 1999; Denness et al., 2011). Lignin is one of the main components of the plant cell wall, acting as an anti-microbial compound by establishing a mechanical barrier, chemically modifying wall-degrading enzymes, or facilitating toxin diffusion, toxic precursor production, and free radical production. The genes involved in lignin production also act as signaling molecules to modulate the defense response (Naoumkina et al., 2010; Wang et al., 2015). Other compounds involved in the plant defense response include flavonoids, monolignols, phenolic acids, stilbenes, phytoalexins, and coumarins (Treutter, 2005; Zaynab et al., 2018). In addition, the phenylpropanoid pathway is associated with the SA signaling pathway. It has been discovered that PAL, a key enzyme in the phenylpropanoid pathway, participates in plant resistance by regulating SA levels (Zhang et al., 2017; Yuan et al., 2019).

B. hispida (wax gourd), the only member of the genus *Benincasa*, is widely cultivated in China and Southeast Asian countries; it is an important crop in the *Cucurbitaceae* family (Ma et al., 2021). One of the most serious threats to *B. hispida* production is wilt disease caused by a *Phytophthora* infection, with *P. melonis* as the primary pathogen (Ren et al., 2020). *P. melonis* falls into Clade 7b, as classified by genus-wide

phylogenetic analyses of *Phytophthora* species (Blair et al., 2008). A *P. melonis* infection causes cucurbit blight; it bursts and spreads quickly during the rainy season, and roots, leaves, stems, and fruits are all susceptible. The underlying resistance mechanisms are still poorly understood, and transcriptome analysis is an effective method for understanding resistance mechanisms. According to the transcriptome analysis of the potato response to P. infestans infection, isolates with different virulence profiles can induce different defense responses at different time points (Duan et al., 2020); in the melon response to P. capsici infection, genes related to plant defense responses are stronger activated in a resistant cultivar than in a susceptible cultivar (Wang et al., 2020). In this study, high-throughput RNA-Seq was used to determine the transcriptome profiles of two *B. hispida* cultivars, B488 (a moderately resistant cultivar) and B214 (a moderately susceptible cultivar), that show significant differences in resistance to *P. melonis*. The identification of the differentially expressed genes in various signal pathways using GO (Gene Ontology) annotations, KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment, and cluster analysis led to the discovery of some genes that may be involved in *B. hispida* resistance to *P. melonis*.

Materials and methods

B. hispida growth, *P. melonis* cultivation

B. hispida cultivars B488 and B214 were provided by the Vegetable Research Institute, Guangdong Academy of Agricultural Sciences. *B. hispida* seedlings were cultivated at 28 °C in a greenhouse under a 16-h light/8-h dark photoperiod for 20 days until the two-leaf stage was reached. The *P. melonis* isolate FS-DY was cultivated on 10% V8 solid medium at 28 °C in the dark. The hyphae were transferred to 10% V8 liquid medium and cultivated 3 days for zoospore harvest.

P. melonis infection

For RNA-seq: A *P. melonis* zoospore suspension (10^5 zoospores per milliliter) was sprayed on the surface of the leaves of 20-day-old B488 and B214 seedlings, sterilized water was used as the control. The *P. melonis* inoculated seedlings and control seedlings were covered, separately. The treated seedlings were cultivated at 28 °C in the dark for 12 hours, then the leaves were dried by filter pepper before harvested in tinfoil and immediately snap-frozen in liquid nitrogen. Three biological replicates were performed for each treatment, and three seedlings for each replicate.

For resistance identification: Five milliliters of a *P. melonis* zoospore suspension (10^3 zoospores per milliliter) was applied to each *B. hispida* seedling. The zoospore suspension was irrigated close to the stem; after irrigation, the seedlings were cultivated at 28 °C under a 16-h light/8-h dark photoperiod. The appearance of browning at the base of a seedling stem and lodging was recorded as death. The survival rate was calculated as the number of survival seedlings divided by the total number of seedlings. Survival rates were determined in triplicate for each cultivar. Ten seedlings were used in each experiment.

RNA sequencing and differentially expressed genes analyses

RNA extraction and sequencing were performed by Shanghai Majorbio Bio-pharm Technology Co., Ltd, Genes Differential Expression and Gene Functional Enrichment Analyses were conducted on the online platform of the Majorbio Cloud Platform (www.majorbio.com). All the sequencing data were deposited in the NCBI Short Read Archive (SRA) database under the BioProject ID: PRJNA895846.

Quantitative real-time RT-PCR

The Bio-Rad Real-time PCR system CFX96 and SYBR Premix (TranStart Green qPCR SuperMix, TransGen Biotech) were used to quantify gene expression. Total RNA extraction (EasyPure Plant RNA Kit, TransGen Biotech) and qRT-PCR were carried out according to the RNA extraction and SYBR Premix instructions, respectively. The specific primer pairs used are listed in Supplementary Table 2. Data were analyzed by the Livak method (Livak and Schmittgen, 2001) and expressed as a normalized relative expression level ($2^{-\Delta\Delta CT}$) of the respective genes, and the relative transcript level of each sample was normalized to actin (Supplementary Table 2). Six biological replicates were included in the experiment. The IBM SPSS version 21 software was used for statistical analysis, the p value was determined using Fisher's protected LSD test.

Results

B488 and B214 had different resistance to *P. melonis*

To select varieties with different resistance to *P. melonis*, we assessed the resistance of 20 *B. hispida* cultivars based on field performance; two of these, B488 and B214, showed significant differences in resistance. The resistance of the two cultivars was reconfirmed using the root-irrigation method. Different concentrations of pathogen could affect the results and plants were hardly survival from a high concentration of pathogen. According to the experiment, we chased a moderate concentration (10^3 zoospores per milliliter) to perform the

root-irrigation experiment and the results showed that B488 is more resistant to *P. melonis* than B214 (Figure 1A), with survival rates of 55 ± 18.0% at 3 dpi, 33.8 ± 6.7% at 4 dpi and 33.8 ± 6.7% at 5 dpi for B448, while B214 had survival rates of 24.4 ± 7.7% at 3 dpi, 0 at 4 dpi and 0 at 5 dpi (Figure 1B). Taken together, these findings suggest that B214 and B488 are respectively a moderately susceptible and resistant cultivar to *P. melonis*.

RNA-seq and quality evaluation

To understand the mechanism causing the differences in the resistance of B488 and B214 to *P. melonis* and the different expression patterns of genes in the early phase of B488 and B214 response to *P. melonis* infection, we performed the inoculation by spraying *P. melonis* zoospore suspension on the surface of the leaves of 20-day-old B488 and B214 seedlings. To ensure the success of infection, a high concentration (10^5 zoospores per milliliter) of *P. melonis* zoospore were chased to perform the experiment. The differentially expressed genes (DEGs) were investigated by RNA sequencing (RNA-seq) 12 h after B488 and B214 were infected with *P. melonis* (PmI). Shanghai Majorbio Bio-pharm Technology Co., Ltd performed the RNA-seq procedure, and the data were analyzed on the Majorbio Cloud Platform online platform (www.majorbio.com).

Pearson correlation and principal component analyses (PCA) (Supplementary Figure 1) were used to compare the samples. The PCA1 had 37% variance and the PCA2 had 18.8% variance, each group was dispersed, and each group sample was gathered. Supplementary Table 1 shows the data from the sequencing of all the samples. The proportion of bases with a quality of no less than 20 after filtration (Q20) was 97-98%, the proportion of bases with a quality of no less than 30 after filtration (Q30) was 92–94%, and the overall data sequencing error rate was less than 0.03%. The base numbers of G and C were 46–47% of the total base number. Mapping rates were greater than 96%. The above results revealed that the RNA-Seq data of these 12 samples were reliable and could be used for subsequent analysis.

DEGs obtained and validation of DEGs by qRT-PCR

The DEGs were identified based on the abundance of RNA-Seq reads and normalized to fragments per kilobase length per million reads (FPKM). A total of 680 (i.e., 638 upregulated and 42 downregulated) and 988 (i.e., 788 upregulated and 200 downregulated) DEGs that showed at least a twofold change in gene expression (q-values < 0.05) were respectively identified in B488 and B214 infected with *P. melonis* infection. A Venn diagram analysis revealed 347 common genes in the DEGs of B488 and B214, as well as 333, 641 distinct genes (Figures 2A, B).

To validate the RNA-Seq results, a quantitative real-time PCR (qRT-PCR) analysis was performed to detect the expression level of nine randomly selected DEGs for B488 and B214. Three genes were randomly selected from the common genes in DEGs of B488 and B214 (*Bhi02G000259*, *Bhi05G000832*, *Bhi07G001503*), three genes from the B488 DEGs of (*Bhi11G001601*, *Bhi01G000575*, *Bhi04G000584*), and three genes from the B214 DEGs of (*Bhi06G000622*, *Bhi11G001731*, *Bhi03G002023*). The qRT-PCR and FPKM data for the nine genes of B488 and B214 showed a similar variation tendency between a mock infection and the PmI (Figure 2C). These findings suggest that the transcriptome data are reliable.

KEGG annotation analysis and GO annotations analysis

An annotation analysis of the B488 and B214 DEGs with KEGG shows that the majority of DEGs are involved in Metabolism, with up to 67.6% and 69.1% for B488 and B214, respectively. Carbohydrate metabolism, biosynthesis of other

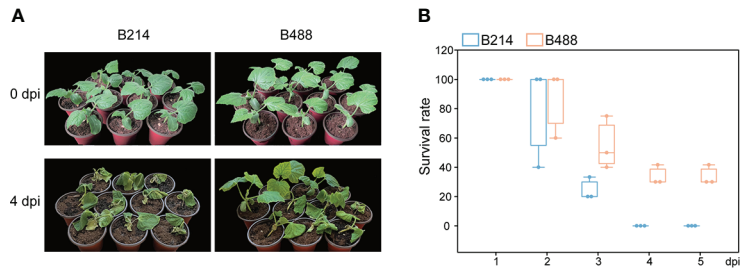


FIGURE 1
P. melonis resistance of B488 and B214. **(A)** Two-leaf stage B488 and B214 seedlings were irrigated with 5 ml of a *P. melonis* zoospore suspension (10^3 zoospores per milliliter) for each seedling, pictures showed the state of B488 and B214 seedlings at 4 days post infection (dpi). **(B)** the survival rates of B488 and B214 were calculated daily for 5 days post infection.

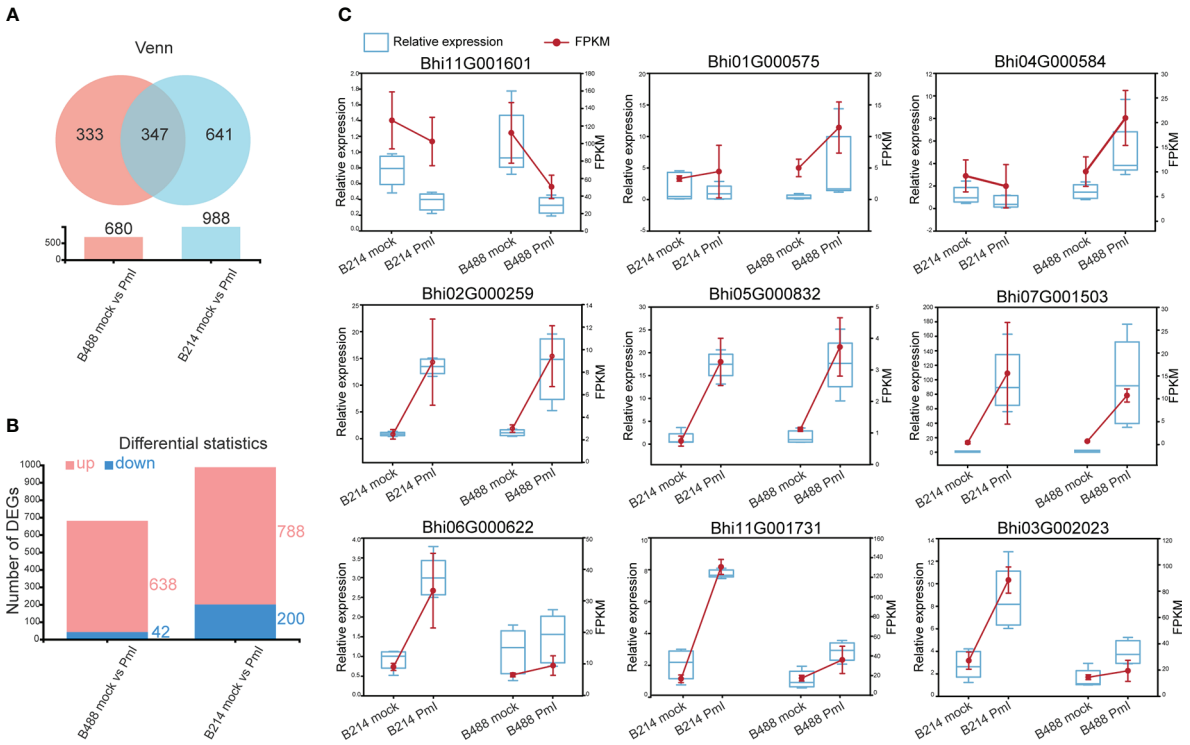


FIGURE 2
DEGs and the validation of selected DEGs using qRT-PCR. **(A)** Venn diagrammatic analysis of DEGs in B488 and B214. **(B)** up- and downregulated DEGs in B488 (mock infection vs. Pmi) and B214 (mock infection vs. Pmi). **(C)** Randomly selected DEGs expression tendency of qRT-PCR data compared with RNA-seq FKPM data.

secondary metabolism, and amino acid metabolism were the pathways with the most enriched genes in the Metabolism category; folding, sorting and degradation, signal transport, transport and catabolism, and environmental adaptation were the pathways with the most enriched genes in the genetic information processing, environmental information processing, cellular processes and organismal systems categories, respectively (Figure 3A).

An analysis of GO annotations revealed that the most relevant terms in the process of *B. hispida* resistance to *P. melonis* were metabolic process and cellular processes in biological process, membrane part and cell part in cellular component, and catalytic activity and binding in molecular function. Furthermore, B214 had more enriched genes than B488 in almost all categories (Figure 3B). The KEGG pathways and GO terms mentioned above may be important in the *B. hispida* response to a *P. melonis* infection.

KEGG enrichment analysis

Phytophthora was defined as a hemibiotrophic pathogen, and the first 18 hours were Biotrophy. In this early phase, hyphae pushed the host cell membrane inwards, forming a direct host-pathogen interface, and the plant cells were not destroyed. This is a critical stage of *Phytophthora* defense. The KEGG enrichment analysis was used to preserve the DEGs involved in this early phase of the defense response. The top four enrichment pathways for the moderately resistant cultivar B488 DEGs were phenylpropanoid biosynthesis (28 genes), plant-pathogen interaction (20 genes), MAPK signaling pathway (17 genes), and plant hormone signal transduction (16 genes). The top four enrichment pathways for the moderately susceptible cultivar B214 DEGs were phenylpropanoid biosynthesis (39 genes), plant-pathogen interaction (25 genes), MAPK signaling pathway (23 genes), and plant hormone signal transduction (20 genes) (Figure 4).

DEGs involved in phenylpropanoid biosynthesis pathway

According to the KEGG enrichment analysis, the DEGs involved in the early phase (i.e., 12 hpi) of the *B. hispida* (i.e., B488 and B214) response to a *P. melonis* infection were enriched in the same primary pathways (i.e., Phenylpropanoid biosynthesis, Plant-pathogen interaction, MAPK signaling

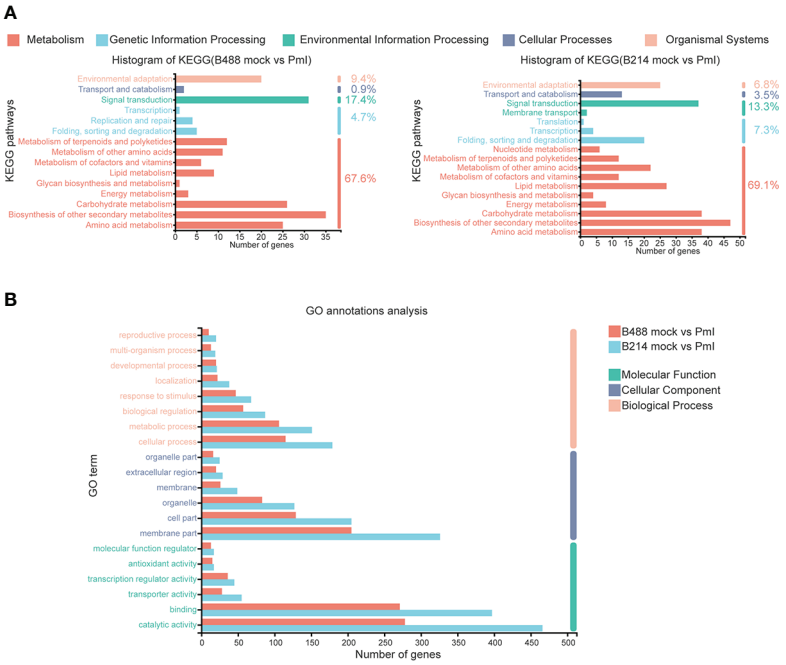


FIGURE 3
KEGG annotations analysis and GO annotations analysis of DEGs. **(A)** KEGG annotations analysis of the DEGs in B488 (mock infection vs. Pml) and B214 (mock infection vs. Pml), the number indicated the percent of gene enrichment in the pathway. **(B)** GO annotations analysis of DEGs in B488 (mock vs. Pml) and B214 (mock vs. Pml).

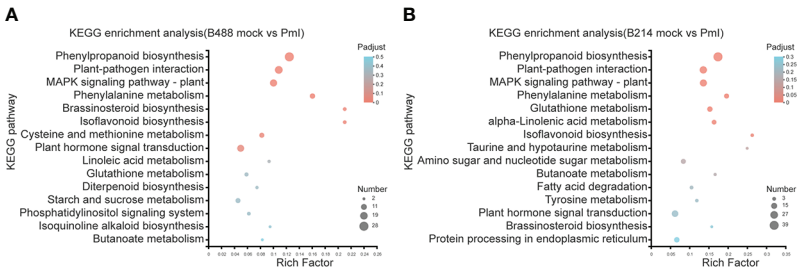


FIGURE 4
KEGG enrichment analysis of DEGs. **(A)** KEGG enrichment analysis of the DEGs in B488. **(B)** KEGG enrichment analysis of the DEGs in B214. The y-axis indicates the pathway names, and the x-axis indicates the rich factor. Rich factor refers to the ratio of the number of DEGs located in the KEGG pathway and the total number of genes in the KEGG pathway. The larger the rich factor the greater the degree of enrichment. The size of the dots indicates the number of genes in the pathway, while the color of the dots indicates the adjust *p*-value ranges.

pathway – plant, and Plant hormone signal transduction). The DEGs were used to perform a cluster analysis to detect the expression profile of the genes in B488 and B214 that were enriched in the aforementioned pathways.

Genes involved in the Phenylpropanoid biosynthesis pathway were clustered into six subclusters based on the expression patterns in the cluster analysis (Supplementary Figure 3). In B488, but not in B214, the *P. melonis* infection induced three genes (Subcluster 4), which coded for peroxidase. In B214, but not in B488, the expression of an oxidoreductase (*Bhi03G000160*) and a transferase gene (*Bhi09G000440*) was reduced (subcluster 5). The phenylalanine ammonia lyase gene (*Bhi10G000432*) was downregulated in B214 but upregulated in B488 (subcluster 6) (Figure 5A; Table 1).

DEGs involved in the plant-pathogen interaction pathway

Plant-pathogen interaction is a pathway that consists of pathogen association receptors, signal transfer molecules, and resistant genes, and is the most important pathway for plant defense. The genes involved in the Plant-pathogen interaction pathway were clustered into six subclusters based on their expression patterns in the cluster analysis. Most genes in this pathway were induced by the *P. melonis* infection, with the exception of four genes in subcluster 5 that were upregulated in B488 but downregulated in B214 (Supplementary Figure 4; Figure 5B). These four genes were predicted to code for two 3-ketoacyl-CoA synthases, a protein kinase, and a basic form of pathogenesis-related protein 1 (Table 2).

DEGs involved in the MAPK signaling pathway–plant

The MAPK cascade is an effective signaling transduction system that plays an important role in plant defense. In the cluster analysis of genes involved in the MAPK signaling pathway-plant (Supplementary Figure 5), most genes were induced by *P. melonis* infection, except for the genes in subcluster 6 (Figure 5C; Table 3). *Bhi05G000499* has been predicted to code for an LRR receptor kinase and may be upstream of the MAPK cascades. Reduction of the expression of *Bhi05G000499* in B214 but not in B488 during a P. melonis infection is an indication that a specific MAPK signaling pathway was more highly activated in B488 than in B214.

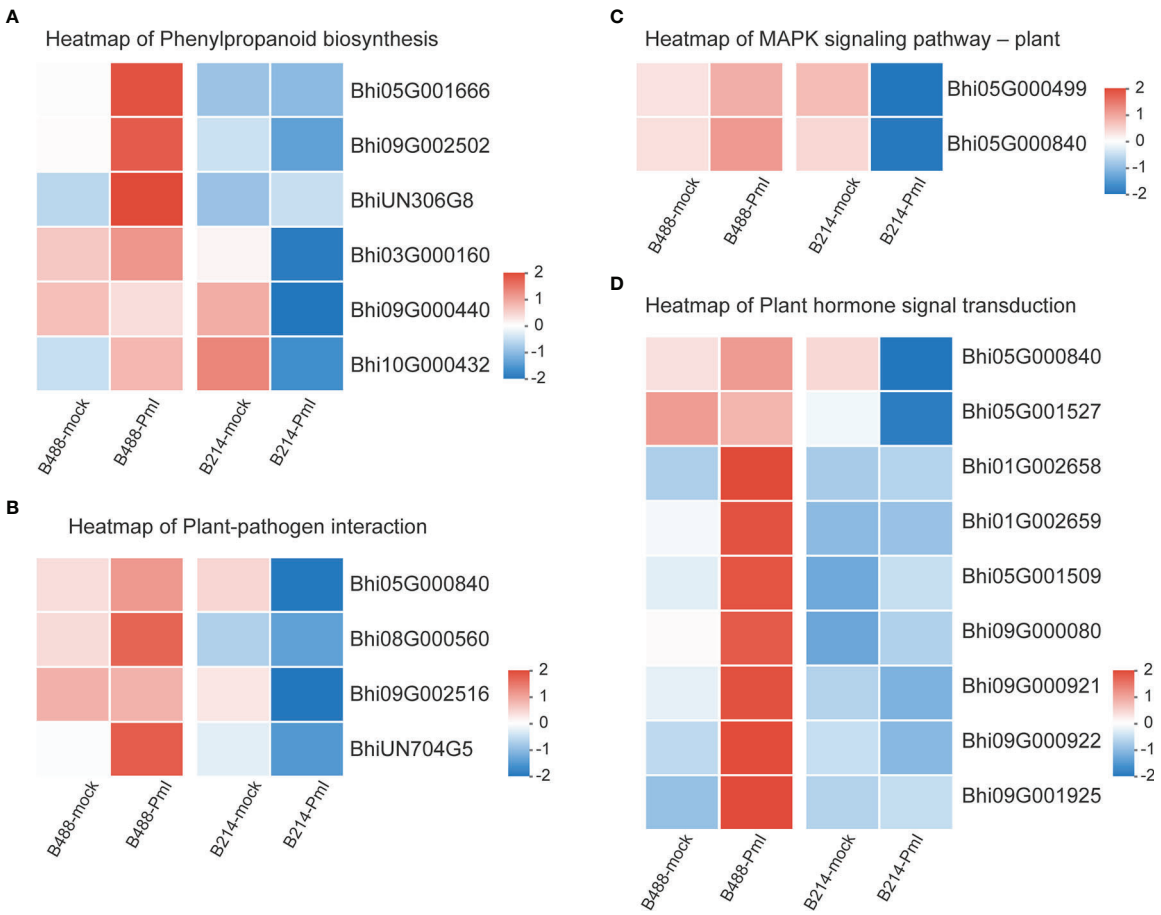


FIGURE 5
Heatmap of DEGs. **(A)** Heatmap of the DEGs in the phenylpropanoid biosynthesis pathway. **(B)** Heatmap of the DEGs in the plant pathogen interaction pathway. **(C)** Heatmap of the DEGs in the MAPK signaling-plant pathway. **(D)** Heatmap of the DEGs in the plant hormone signal transduction pathway. Log_{10}(FPKM+1) values were used in the heatmap, and the average values of FPKM from the three biological replicates were used for log_{10} (FPKM+1) calculation. Red indicates high expression, and blue indicates low expression.

TABLE 1 Genes in subcluster 4, 5 and 6 of phenylpropanoid biosynthesis pathway.

subcluster	Gene id	NR description
4	*Bhi05G001666*	peroxidase P7-like
	Bhi09G002502	peroxidase 18
	BhiUN306G8	lignin-forming anionic peroxidase-like
5	*Bhi03G000160*	flavin-dependent oxidoreductase FOX2-like
	Bhi09G000440	quinate hydroxycinnamoyl transferase
6	*Bhi10G000432*	phenylalanine ammonia lyase-like
NR (Non-Redundant Protein Sequence Database).		

TABLE 2 Genes in subcluster 5 of plant pathogen interaction pathway.

subcluster	Gene ID	NR description
5	*Bhi08G000560*	3-ketoacyl-CoA synthase 12-like
	BhiUN704G5	calcium-dependent protein kinase 29
	Bhi05G000840	basic form of pathogenesis-related protein 1-like
	Bhi09G002516	3-ketoacyl-CoA synthase 10
NR (Non-Redundant Protein Sequence Database).		

DEGs involved in plant hormone signal transduction pathway

To identify the hormones that participate in *B. hispida* response to *P. melonis* infection, a cluster analysis of genes involved in Plant hormone signal transduction was performed (Supplementary Figure 6). The results show that genes in subcluster 4 were downregulated by a *P. melonis* infection of B214 but not B488, and genes in subcluster 6 were upregulated in B488 but not in B214 (Figure 5D). With the exception of the *Bhi05G000840* gene, which was predicted to code for a basic form of pathogenesis-related protein 1-like and was related to the SA (salicylic acid) signaling pathway, the others were all involved in the auxin signaling pathway (Table 4), and the gene expression tendency led to a similar result; i.e., more auxin-induced/responsive protein was accumulated in B488 than B214 at 12 hpi, which suggests that auxin signaling is a key signaling pathway that contributes to the difference in resistance between B488 and B214 to a *P. melonis* infection.

Transcription factor genes Involved in *B. hispida* response to *P. melonis* infection

It has been suggested that transcription factors (TFs) can play either a positive or negative role in the plant defense response. A total of 51 and 78 genes, respectively, of the DEGs of B488 and B214 are predicted to encode TF proteins. A total of 51 TF genes of the DEGs of B488 were categorized into 9 TF families according to statistics on the TF families (Figure 6A); the top three were the WRKY, ERF, and MYB families with 18, 9, and 8 members, respectively. The 78 TF genes in the B214 DEGs were classified into 19 TF families (Figure 6B), with the primary TF families of WRKY (19 genes), NAC (10 genes), and MYB (10 genes). A cluster analysis of TF genes involved in the B. hispida response to a P. melonis infection revealed that genes in subclusters 5 and 6 have different expression patterns between B488 and B214, with increased expression in B488 and decreased expression in B214 (Supplementary Figure 7; Figure 6C) (Table 5).

TABLE 3 Genes in subcluster 6 of MAPK signaling pathway−plant.

subcluster	Gene ID	NR description
6	*Bhi05G000840*	basic form of pathogenesis-related protein 1-like
	Bhi05G000499	LRR receptor-like serine/threonine-protein kinase ERL1
NR (Non-Redundant Protein Sequence Database).		

TABLE 4 Genes in subcluster 4 and 6 of plant hormone signal transduction pathway.

subcluster	Gene id	NR description
4	*Bhi05G000840*	basic form of pathogenesis-related protein 1-like
	Bhi05G001527	auxin-responsive protein SAUR50-like
6	*Bhi01G002658*	auxin-induced protein 22A-like
	Bhi01G002659	auxin-induced protein AUX22-like
	Bhi05G001509	auxin-responsive protein
	Bhi09G000080	auxin-induced protein AUX22-like
	Bhi09G000921	auxin-induced protein 22D-like
	Bhi09G000922	auxin-induced protein AUX28-like
	Bhi09G001925	auxin-responsive protein IAA29-like

NR (Non-Redundant Protein Sequence Database).

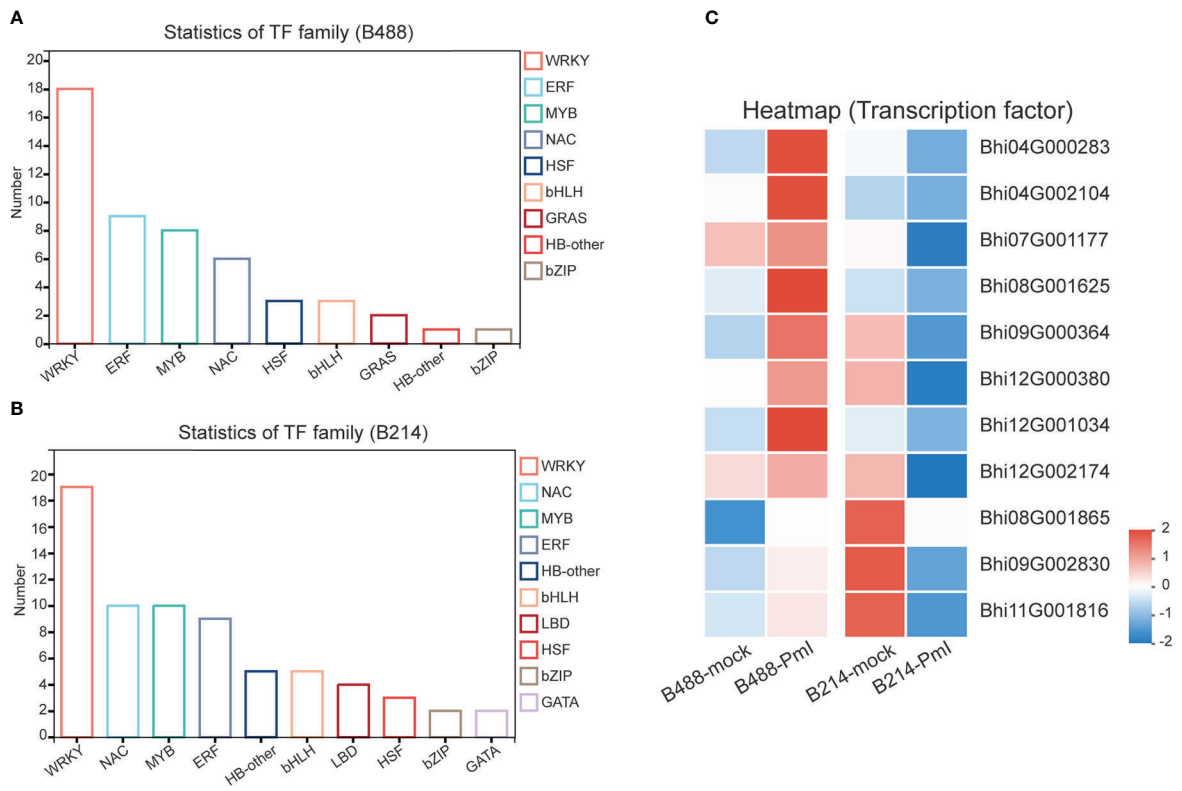


FIGURE 6
TF family of DEGs. **(A)** Statistics for TF in B488 (mock infection vs. Pml). **(B)** Statistics of TF in B214 (mock infection vs. Pml). **(C)** Heatmap of differentially expressed TF genes between B488 (mock infection vs. Pml) and B214 (mock infection vs. Pml).

Detection of gene expression at 6 hpi and 12 hpi by qRT-PCR, allowed validation of the expression tendency of TF genes in B488 and B214 and identification of the TF genes induced by a *P. melonis* infection in the early phase. Six genes were significantly upregulated in B488 but were not significantly different in B214 (Figure 7), of which four (i.e., *Bhi04G000283*, *Bhi08G001625*, *Bhi09G000364* and *Bhi12G001034*) were upregulated at 6 hpi and 12 hpi and were predicted to code

TABLE 5 Genes in subcluster 5 and 6 of TF genes.

subcluster	Gene id	NR description
5	*Bhi04G000283*	transcription factor MYB44-like
	Bhi04G002104	ethylene-responsive transcription factor ERF020
	Bhi07G001177	transcription factor bHLH49 isoform X3
	Bhi08G001625	NAC domain-containing protein 86
	Bhi09G000364	GATA transcription factor 9-like
	Bhi12G000380	axial regulator YABBY 5-like
	Bhi12G001034	NAC domain-containing protein 45
	Bhi12G002174	squamosa promoter-binding-like protein 16
6	*Bhi08G001865*	ethylene-responsive transcription factor 11
	Bhi09G002830	ethylene-responsive transcription factor 4-like
	Bhi11G001816	floral homeotic protein APETALA 2-like

NR (Non-Redundant Protein Sequence Database).

for MYB transcription factor, NAC transcription factor, GATA transcription factor and NAC transcription factor, respectively (Table 5). These four genes could be candidate genes that contribute to the different levels of resistance to *P. melonis* of B488 and B214.

Discussion

Benincasa hispida is an important crop of the *Cucurbitaceae* family, and currently available cultivars are particularly vulnerable to *Phytophthora melonis*, one of the principal threats to wax gourd production. In this study, high throughput sequencing of two *B hispida* cultivars, B488 (a moderately resistant cultivar) and B214 (a moderately susceptible cultivar) with different degrees of resistance to *P melonis*, was performed to elucidate the mechanism of *B hispida* resistance to *P melonis* and identify the genes involved in the defense response. In comparison to a mock infection, 680 and 988 DEGs were respectively found in the presence of a *P. melonis* infection with B488 and B214 (Figure 2A). Phenylpropanoid biosynthesis, Plant-pathogen interaction, MAPK signaling pathway–plant, and Plant hormone signal transduction were discovered as the most active pathways when *B hispida* is

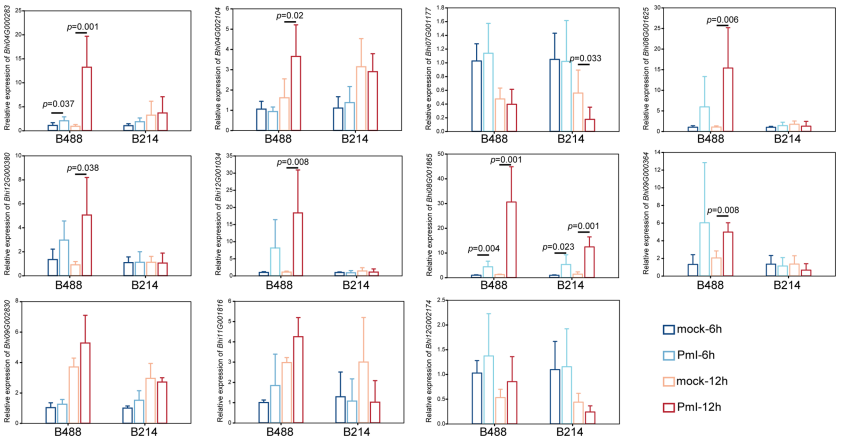


FIGURE 7
Expression pattern of TF genes. A *P. melonis* zoospore suspension (10^5 zoospores per milliliter) was sprayed on surface of B488 and B214 leaves, sterilized water was sprayed on surface of B488 and B214 leaves as the control; leaves were harvested at 6 h and 12h for RNA extraction and subsequent qRT-PCR analysis, six biological replicates were performed, *p* value was determined using Fisher's protected LSD test.

attacked by *P melonis*. Comparing the expression patterns of the DEGs involved in the aforementioned pathways in B488 and B214 suggests some candidate genes that could contribute to the differences in the resistance of B488 and B214.

Phenylpropanoids are secondary metabolites that plants produce to improve their development and cope with biotic and abiotic stresses (Tohge et al., 2017). According to a KEGG enrichment analysis of DEGs at an early stage of a *P. melonis* infection of *B. hispida*, we discovered that phenylpropanoid biosynthesis is the most frequently seen enrichment pathway, and almost all the DEGs in this pathway are upregulated (Figure S3). One gene (*Bhi10G000432*), which was predicted to code phenylalanine ammonia lyase (PAL), was consistently upregulated in B488 and downregulated in B214 (Figure 5A). PAL is the gateway enzyme of the general phenylpropanoid pathway (Zhang and Liu, 2015); it has direct control over downstream metabolites such as lignin and flavonoid biosynthesis, which are linked to plant resistance. PAL can also contribute to plant resistance by regulating the SA levels (Zhang et al., 2017; Yuan et al., 2019). Direct evidence suggests that the overexpression of PAL will enhance the resistance and knockdown or silencing of PAL will increase the susceptibility of the plant to pathogens (Yadav et al., 2020). The higher expression of PAL in B488 than in B214 is associated with the higher resistance of the former, and the mechanism requires further investigation.

MAPK cascades are involved in a variety of defense events, including defense gene expression, phytoalexin and defense hormone biosynthesis, and stomatal immunity (Liu and Zhang, 2004; Li et al., 2012; Su et al., 2017). MAPKKKs are phosphorylated directly by PRRs or by receptor-like cytoplasmic kinases (RLCKs) that act downstream of PRRs in MAPK cascades, and evidence suggests that the same MAPKKK acting downstream of different PRRs may result in a different MAPK cascade activation (Yamada et al., 2016). Our study indicated a gene (*Bhi05G000499*) predicted to code for an LRR receptor kinase that is induced by *P. melonis* infection in B488 (Figure 5C). It is reasonable to hypothesize that this LRR receptor kinase activates a MAPK cascade in B488 that differs from the MAPK cascades activated in B214, resulting in the greater expression of genes involved in the Plant-pathogen interaction pathway (i.e., *Bhi08G000560*, *BhiUN704G5*, *Bhi05G000840*, *Bhi09G002516*) (Figure 5B).

Auxin, an important phytohormone, is involved in plant growth, development, and environmental stimuli, as well as plant-microbe interactions (Benjamins and Scheres, 2008). It is thought to increase susceptibility to the establishment of mutualistic or pathogenic microbes in plant tissues in plant-microbe interactions (Kunkel and Harper, 2018; McClerklin et al., 2018). In a study of soybean response to *Phytophthora sojae*, levels of auxin and related metabolites were significantly increased in soybean at 48 and 72 hpi, and increased auxin levels have been proposed to improve soybean susceptibility (Stasko et al., 2020). In this study, eight genes (i.e., *Bhi05G001527*, *Bhi01G002658*, *Bhi01G002659*, *Bhi05G001509*, *Bhi09G000080*, *Bhi09G000921*, *Bhi09G000922*, *Bhi09G001925*), predicted to code the AUX/IAA protein show different expression patterns in B488 and B214 after infection with *P melonis*, and it is believed that the Aux/IAA proteins are transcriptional co-regulators that function as repressors of the early auxin response genes at low auxin concentrations (Weijers and Wagner, 2016). In our study, the expressions of *Bhi01G002658*, *Bhi01G002659*, higher in the moderately resistant cultivar (B488), but not in the moderately susceptible cultivar (B214), reveal a stronger auxin signaling repression in B488 than in B214 during the early phase of *P melonis* infection. Furthermore, two indole-3-acetic acid-amido synthetase genes (i.e., *Bhi03G000746* and *Bhi09G001660*) were upregulated in B214 but not in B488 (Figure S6), implying that B214 accumulates more auxin than B488. Collectively, the auxin signaling repressor and auxin concentration are critical in the of the resistance *B hispida* to *P melonis*.

Many genes are reprogrammed during the defense response process, which is dependent on transcription factor regulation. Most TFs bind to a specific DNA sequence in the promoter region, and either activate or repress gene expression (Weirauch et al., 2014). A strategy for identifying TFs that function in plant defense is to first identify TF genes that show altered transcription levels during the initial period of the defense response. In this study, we found four TF genes are upregulated at 6 h and 12 h in B488 but not in B214 after infection with *P melonis*. *Bhi04G000283* encodes a MYB (myeloblastosis related) transcription factor, while *Bhi08G001625*, *Bhi12G001034*, and *Bhi09G000364* encode NAC [no apical meristem (NAM), Arabidopsis transcription activation factor (ATAF1/2), and cup-shaped cotyledon (CUC2)] transcription factor, NAC transcription factor and GATA transcription factor, respectively. MYB and NAC are two major TF families involved in plant defense (Ng et al., 2018). Indeed, MYB transcription factors participate in the defense response by modulating the biosynthesis of secondary metabolites (Wang et al., 2021). For instance, MYB34/51/122 contribute to resistance toward *Plectosphaerella cucumerina* exclusively through indolic glucosinolate biosynthesis (Frerigmann et al., 2016). NAC transcription factor positively regulates disease resistance by suppressing the ABA signaling pathway (Liu et al., 2018).GATA factors are evolutionarily conserved transcription factors that are found in animals, fungi, and plants. Previous studies of the biological roles of plant GATAs revealed that their major functions are associated with plant development (Schwechheimer et al., 2022), Nevertheless, we found that *Bhi09G000364* may be involved in the resistance to *P melonis* in *B hispida*, which is a new biological role for GATA factors

This study is the first report of the identification of genes that contribute to the defense of *B hispida* to *P melonis*. A COG (Clusters of Orthologous Groups) classification of DEGs in B488 and B214 shows that 59% and 55% of the genes, respectively, have

unknown functions (Supplementary Figure 2), suggesting a great potential for the identification of *B hispida* resistance-associated genes. A subsequent functional analysis of these candidate genes could reveal the mechanism of the defense of *B hispida* against *P melonis* and identify molecular markers for developing *B hispida* cultivars with high levels resistance to *P melonis*.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number: PRJNA895846.

Author contributions

JC, SY, and DX conceived and designed the experiments. JC and JY performed the experiments. JC, WL, and DX analyzed the data. JC wrote the manuscript. SY, WL, JY, BJ, and DX reviewed and revised the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This research was funded by Laboratory of Lingnan Modern Agriculture Project (NT2021004), Agricultural Competitive Industry Discipline Team Building Project of Guangdong Academy of Agricultural Sciences (202103TD), the Department of Agriculture and Rural Areas of Guangdong Province of China (2022KJ110).

Acknowledgments

The authors thank the Plant Protection Research Institute Guangdong Academy of Agricultural Sciences for supplying the *P. melonis* isolate (FS-DY).

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.2022.1106123/full#supplementary-material

References

Aerts, N., Pereira Mendes, M., and Van Wees, S. C. M. (2021). Multiple levels of crosstalk in hormone networks regulating plant defense. *Plant J.* 105, 489–504. doi: 10.1111/tpj.15124

Appiah, A. A., van West, P., Osborne, M. C., and Gow, N. A. (2005). Potassium homeostasis influences the locomotion and encystment of zoospores of plant pathogenic oomycetes. *Fungal Genet. Biol.* 42, 213–223. doi: 10.1016/j.fgb.2004.11.003

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in arabidopsis innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a

Bechinger, C., Giebel, K. F., Schnell, M., Leiderer, P., Deising, H. B., and Bastmeyer, M. (1999). Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus. *Science* 285, 1896–1899. doi: 10.1126/science.285.5435.1896

Benjamins, R., and Scheres, B. (2008). Auxin: the looping star in plant development. *Annu. Rev. Plant Biol.* 59, 443–465. doi: 10.1146/annurev.arplant.58.032806.103805

Bi, G., Zhou, Z., Wang, W., Li, L., Rao, S., Wu, Y., et al. (2018). Receptor-like cytoplasmic kinases directly link diverse pattern recognition receptors to the activation of mitogen-activated protein kinase cascades in arabidopsis. *Plant Cell* 30, 1543–1561. doi: 10.1105/tpc.17.00981

Blair, J. E., Coffey, M. D., Park, S. Y., Geiser, D. M., and Kang, S. (2008). A multi-locus phylogeny for phytophthora utilizing markers derived from complete genome sequences. *Fungal Genet. Biol.* 45, 266–277. doi: 10.1016/j.fgb.2007.10.010

Chialva, M., Lanfranco, L., and Bonfante, P. (2022). The plant microbiota: composition, functions, and engineering. *Curr. Opin. Biotechnol.* 73, 135–142. doi: 10.1016/j.copbio.2021.07.003

Denness, L., McKenna, J. F., Segonzac, C., Wormit, A., Madhou, P., Bennett, M., et al. (2011). Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in arabidopsis. *Plant Physiol.* 156, 1364–1374. doi: 10.1104/pp.111.175737

Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y. , Li, X., et al. (2018). Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity. *Cell* 173, 1454–1467 e15. doi: 10.1016/j.cell.2018.03.044

Duan, Y., Duan, S., Armstrong, M. R., Xu, J., Zheng, J., Hu, J., et al. (2020). Comparative transcriptome profiling reveals compatible and incompatible patterns of potato toward phytophthora infestans. *G3 (Bethesda)* 10, 623–634. doi: 10.1534/g3.119.400818

Frerigmann, H., Pislewska-Bednarek, M., Sanchez-Vallet, A., Molina, A., Glawischnig, E., Gigolashvili, T., et al. (2016). Regulation of pathogen-triggered tryptophan metabolism in arabidopsis thaliana by MYB transcription factors and indole glucosinolate conversion products. *Mol. Plant* 9, 682–695. doi: 10.1016/j.molp.2016.01.006

Gao, M., Liu, J., Bi, D., Zhang, Z., Cheng, F., Chen, S., et al. (2008). MEKK1, MKK1/MKK2 and MPK4 function together in a mitogen-activated protein kinase cascade to regulate innate immunity in plants. *Cell Res.* 18, 1190–1198. doi: 10.1038/cr.2008.300

Huang, W., Wang, Y., Li, X., and Zhang, Y. (2020). Biosynthesis and regulation of salicylic acid and n-hydroxypipecolic acid in plant immunity. *Mol. Plant* 13, 31–41. doi: 10.1016/j.molp.2019.12.008

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Jung, T., Horta Jung, M., Webber, J. F., Kageyama, K., Hieno, A., Masuya, H., et al. (2021). The destructive tree pathogen phytophthora ramorum originates from the laurosilva forests of East Asia. *J. Fungi (Basel)* 7, 226. doi: 10.3390/jof7030226

Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., et al. (2015). The top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* 16, 413–434. doi: 10.1111/mpp.12190

Kourelis, J., and van der Hoorn, R. A. L. (2018). Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for r protein function. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579

Kunkel, B. N., and Harper, C. P. (2018). The roles of auxin during interactions between bacterial plant pathogens and their hosts. *J. Exp. Bot.* 69, 245–254. doi: 10.1093/jxb/erx447

Lamour, K. H., Stam, R., Jupe, J., and Huitema, E. (2012). The oomycete broad-host-range pathogen phytophthora capsici. *Mol. Plant Pathol.* 13, 329–337. doi: 10.1111/j.1364-3703.2011.00754.x

Li, G., Meng, X., Wang, R., Mao, G., Han, L., Liu, Y., et al. (2012). Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in arabidopsis. *PloS Genet.* 8, e1002767. doi: 10.1371/journal.pgen.1002767

Liu, Q., Yan, S., Huang, W., Yang, J., Dong, J., Zhang, S., et al. (2018). NAC transcription factor ONAC066 positively regulates disease resistance by suppressing the ABA signaling pathway in rice. *Plant Mol. Biol.* 98, 289–302. doi: 10.1007/s11103-018-0768-z

Liu, Y., and Zhang, S. (2004). Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in arabidopsis. *Plant Cell* 16, 3386–3399. doi: 10.1105/tpc.104.026609

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

Ma, L., Liu, Z., Cheng, Z., Gou, J., Chen, J., Yu, W., et al. (2021). Identification and application of BhAPRR2 controlling peel colour in wax gourd (Benincasa hispida). *Front. Plant Sci.* 12, 716772. doi: 10.3389/fpls.2021.716772

McClerklin, S. A., Lee, S. G., Harper, C. P., Nwumeh, R., Jez, J. M., and Kunkel, B. N. (2018). Indole-3-acetaldehyde dehydrogenase-dependent auxin synthesis contributes to virulence of pseudomonas syringae strain DC3000. *PloS Pathog.* 14, e1006811. doi: 10.1371/journal.ppat.1006811

Mora-Sala, B., Leon, M., Perez-Sierra, A., and Abad-Campos, P. (2022). New reports of phytophthora species in plant nurseries in Spain. *Pathogens* 11, 826. doi: 10.3390/pathogens11080826

Naoumkina, M. A., Zhao, Q., Gallego-Giraldo, L., Dai, X., Zhao, P. X., and Dixon, R. A. (2010). Genome-wide analysis of phenylpropanoid defence pathways. *Mol. Plant Pathol.* 11, 829–846. doi: 10.1111/j.1364-3703.2010.00648.x

Ng, D. W., Abeysinghe, J. K., and Kamali, M. (2018). Regulating the regulators: The control of transcription factors in plant defense signaling. *Int. J. Mol. Sci.* 19, 3737. doi: 10.3390/ijms19123737

Ren, R., Yang, X., Song, A., Li, C., Yang, H., and Kang, Y. (2020). Control of phytophthora melonis damping-off treated with 24-epibrassinolide and a histological study of cucumber hypocotyl. *Protoplasma* 257, 1519–1529. doi: 10.1007/s00709-020-01523-y

Schwechheimer, C., Schroder, P. M., and Blaby-Haas, C. E. (2022). Plant GATA factors: Their biology, phylogeny, and phylogenomics. *Annu. Rev. Plant Biol.* 73, 123–148. doi: 10.1146/annurev-arplant-072221-092913

Stasko, A. K., Batnini, A., Bolanos-Carriel, C., Lin, J. E., Lin, Y., Blakeslee, J. J., et al. (2020). Auxin profiling and GmPIN expression in phytophthora sojae-soybean root interactions. *Phytopathology* 110, 1988–2002. doi: 10.1094/PHYTO-02-20-0046-R

Suarez-Rodriguez, M. C., Adams-Phillips, L., Liu, Y., Wang, H., Su, S. H., Jester, P. J., et al. (2007). MEKK1 is required for flg22-induced MPK4 activation in arabidopsis plants. *Plant Physiol.* 143, 661–669. doi: 10.1104/pp.106.091389

Su, J., Zhang, M., Zhang, L., Sun, T., Liu, Y., Lukowitz, W., et al. (2017). Regulation of stomatal immunity by interdependent functions of a pathogen-responsive MPK3/MPK6 cascade and abscisic acid. *Plant Cell* 29, 526–542. doi: 10.1105/tpc.16.00577

Tang, D., Wang, G., and Zhou, J. M. (2017). Receptor kinases in plant-pathogen interactions: More than pattern recognition. *Plant Cell* 29, 618–637. doi: 10.1105/tpc.16.00891

Tian, W., Hou, C., Ren, Z., Wang, C., Zhao, F., Dahlbeck, D., et al. (2019). A calmodulin-gated calcium channel links pathogen patterns to plant immunity. *Nature* 572, 131–135. doi: 10.1038/s41586-019-1413-y

Tohge, T., de Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi: 10.1093/jxb/erx177

Treutter, D. (2005). Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biol. (Stuttg)* 7, 581–591. doi: 10.1055/s-2005-873009

Van de Weyer, A. L., Monteiro, F., Furzer, O. J., Nishimura, M. T., Cevik, V. , Witek, K., et al. (2019). A species-wide inventory of NLR genes and alleles in arabidopsis thaliana. *Cell* 178, 1260–1272 e14. doi: 10.1016/j.cell.2019.07.038

Wang, G. F., He, Y., Strauch, R., Olukolu, B. A., Nielsen, D., Li, X., et al. (2015). Maize homologs of hydroxycinnamoyltransferase, a key enzyme in lignin biosynthesis, bind the nucleotide binding leucine-rich repeat Rp1 proteins to modulate the defense response. *Plant Physiol.* 169, 2230–2243. doi: 10.1104/pp.15.00703

Wang, X., Niu, Y., and Zheng, Y. (2021). Multiple functions of MYB transcription factors in abiotic stress responses. *Int. J. Mol. Sci.* 22, 6125. doi: 10.3390/ijms22116125

Wang, P., Wu, H., Zhao, G., He, Y., Kong, W., Zhang, J., et al. (2020). Transcriptome analysis clarified genes involved in resistance to phytophthora capsici in melon. *PloS One* 15, e0227284. doi: 10.1371/journal.pone.0227284

Wasternack, C., and Song, S. (2017). Jasmonates: biosynthesis, metabolism, and signaling by proteins activating and repressing transcription. *J. Exp. Bot.* 68, 1303–1321.

Weijers, D., and Wagner, D. (2016). Transcriptional responses to the auxin hormone. *Annu. Rev. Plant Biol.* 67, 539–574. doi: 10.1146/annurev-arplant-043015-112122

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. doi: 10.1016/j.cell.2014.08.009

Yadav, V., Wang, Z., Wei, C., Amo, A., Ahmed, B., Yang, X., et al. (2020). Phenylpropanoid pathway engineering: An emerging approach towards plant defense. *Pathogens* 9, 312. doi: 10.3390/pathogens9040312

Yamada, K., Yamaguchi, K., Shirakawa, T., Nakagami, H., Mine, A., Ishikawa, K., et al. (2016). The arabidopsis CERK1-associated kinase PBL27 connects chitin perception to MAPK activation. *EMBO J.* 35, 2468–2483. doi: 10.15252/embj.201694248

Yuan, W., Jiang, T., Du, K., Chen, H., Cao, Y., Xie, J., et al. (2019). Maize phenylalanine ammonia-lyases contribute to resistance to sugarcane mosaic virus infection, most likely through positive regulation of salicylic acid accumulation. *Mol. Plant Pathol.* 20, 1365–1378. doi: 10.1111/mpp.12817

Zaynab, M., Fatima, M., Abbas, S., Sharif, Y., Umair, M., Zafar, M. H., et al. (2018). Role of secondary metabolites in plant defense against pathogens. *Microb. Pathog.* 124, 198–202. doi: 10.1016/j.micpath.2018.08.034

Zhang, Y., and Li, X. (2019). Salicylic acid: biosynthesis, perception, and contributions to plant immunity. *Curr. Opin. Plant Biol.* 50, 29–36. doi: 10.1016/j.pbi.2019.02.004

Zhang, X., and Liu, C. J. (2015). Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol. Plant* 8, 17–27. doi: 10.1016/j.molp.2014.11.001

Zhang, C., Wang, X., Zhang, F., Dong, L., Wu, J., Cheng, Q., et al. (2017). Phenylalanine ammonia-lyase2.1 contributes to the soybean response towards phytophthora sojae infection. *Sci. Rep.* 7, 7242. doi: 10.1038/s41598-017-07832-2

Zhou, J. M., and Zhang, Y. (2020). Plant immunity: Danger perception and signaling. *Cell* 181, 978–989. doi: 10.1016/j.cell.2020.04.028

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fpls.2023.1127032

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Shuping Qu,
Northeast Agricultural University, China
Jiang Shi,
Tea Research Institute, Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Junxing Li
lijunxing@gdaas.cn

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

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

RECEIVED 19 December 2022
ACCEPTED 11 January 2023
PUBLISHED 26 January 2023

CITATION
Deng L, Yang X, Qiu Y, Luo J, Wu H, Liu X, Zhao G, Gong H, Zheng X and Li J (2023) Metabolic and molecular mechanisms underlying the foliar Zn application induced increase of 2-acetyl-1-pyrroline conferring the 'taro-like' aroma in pumpkin leaves.
Front. Plant Sci. 14:1127032.
doi: 10.3389/fpls.2023.1127032

Metabolic and molecular mechanisms underlying the foliar Zn application induced increase of 2-acetyl-1-pyrroline conferring the 'taro-like' aroma in pumpkin leaves

Liting Deng[1†], Xian Yang[2†], Yuehan Qiu[1,2], Jianning Luo[1], Haibin Wu[1], Xiaoxi Liu[1], Gangjun Zhao[1], Hao Gong[1], Xiaoming Zheng[1] and Junxing Li[1*]

[1]Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China, [2]College of Horticulture, South China Agricultural University, Guangzhou, Guangdong, China

Introduction: Fresh pumpkin leaf is popular vegetable for its rich nutrition. The pleasant taro-like odour is important aroma quality of crops, and mostly contributed by 2-acetyl-1-pyrroline in pumpkin. Element Zn can impact metabolite biosynthesis in plants, including aroma formation. However, Zn-induced biochemical responses, especially 2-acetyl-1-pyrroline formation in pumpkin, haven't been elucidated.

Methods: This study integrated metabolome and transcriptome to explore molecular fluctuations in pumpkin leaves at different time intervals after foliar Zn treatment.

Result and Discussion: We first identified more than one thousand metabolites from pumpkin leaves by integrating different mass spectrometry methods according to the form in which a metabolite exists. Comparative metabolomic analysis revealed there were separately 25 out of 50 and 286 out of 963 metabolites that were respectively identified by gas chromatography-mass spectrometry and liquid chromatography-tandem mass spectrometry, differentially regulated by Zn treatment. Our findings revealed that 50mg/L of Zn significantly enhanced 2-acetyl-1-pyrroline production by more than 38%, which was contributed by increased biosynthesis of its precursors, including ornithine and proline. The following transcriptome analysis discovered 30,574 genes, including 953 novel genes. Zn treatment induced the differential expression of 41.6% of identified genes which were supposed to regulate the downstream metabolite changes in a time-dependent manner. Pathway analysis indicated that alternations in primary metabolism, including carbon metabolism and biosynthesis of amino acids, were vital to the fluctuated aromatic compound generation. Phytohormones and transcription factors may regulate the expression of gene *P5CS* and proline biosynthesis, which, therefore, affect 2-acetyl-1-pyrroline production. This research reveals molecular mechanisms of 2-acetyl-

1-pyrroline formation in pumpkin, which will provide the molecular basis for desired aroma compound production through metabolite engineering.

KEYWORDS

2-acetyl-1-pyrroline, aroma, pumpkin, metabolome, transcriptome, Zn

1 Introduction

Aroma compounds are one of the biggest classes of metabolites, and important factors that impact the character and quality of food, and economic value, including vegetables, fruits, and beverages (Zeng et al., 2019). Only a few key odorants with significant concentration and exceeding their odour thresholds can predominantly contribute to the overall aroma profiles of plants, and therefore be detected by the human olfactory system (Poehlmann and Schieberle, 2013). Therefore, the characteristic aroma component, especially those that act as the principal contributors to the pleasant odour, becomes the vital agronomic and economic trait of plants. Cucurbitaceae plants are the species that provide most food for human consumption, among which pumpkin is a well-known source for the accumulation of different metabolites and valuable nutrients (Mashiane et al., 2021). Except for fruits, pumpkin leaves are believed to be safe and nutritious vegetables, and with great popularity in diets of Asian and African consumers (Mashiane et al., 2021). Consumer liking is strongly correlated with aromatic compounds. Pumpkin leaves generate various metabolites, and produce rich aroma. On the market, pumpkin varieties with distinctive taro-like flavour are more preferred. The successful and accurate distinguishment of different pumpkin varieties with and without taro-like odour sensing from either leaves or fruits by the characteristic component 2-AP proved it is the key contributor of the distinctive odour (Li et al., 2019; Li et al., 2022). However, mechanisms related to 2-AP production and accumulation in pumpkin are not clearly covered. Additionally, the involvement of both internal genetic factors and external environmental changes in the regulation of 2-AP formation increases the complexcity of associated mechanisms.

In nature, metabolites generation is a biochemical and physiological response to environmental stresses that plants derived during the evolutionary process. It is interesting that these generated metabolites from stress responses can be harmful to plants or potentially improve the important quality components in crops (Zeng et al., 2019). For instance, stress application to leaves impacts the aroma biosynthesis and tea quality (Zeng et al., 2019). Hence, artificially applying safe and effective stimulates is of great importance to enhance the production of the characteristic compounds and improve aroma quality of vegetables. In practice, exogenous application of micronutrients at low concentration can change the aroma formation, therefore improving the aromatic characteristics of plants (Bao et al., 2021). For instance, increased number of volatile metabolites and their total concentrations were observed in grape after foliar Zinc sulfate ($ZnSO_4$) treatment (Song et al., 2016). Zinc chloride ($ZnCl_2$) application positively impacted the yield, together with the aroma of fragrant rice, especially the accumulation of key aroma contributor 2-AP (Bao et al., 2021). It is well-known that Zn is an essential microelement in plants, and involves in the regulation of diverse physiological and biochemical processes at all developmental and metabolic stages (Broadley et al., 2007). It is reported that Zn-mediated changes in carbohydrate metabolism and biosynthesis of plant hormones were involved in the regulation of 2-AP production in fragrant rice (Bao et al., 2021). Potentially, exogenous application of Zn could impact the aroma production in pumpkin. However, the effect of Zn element on the 2-AP production in pumpkin, and the underlying regulatory mechanisms has not been demonstrated yet.

It is suggested that 2-AP could be initially synthesised in leaves of the fragrant rice (Hinge et al., 2016; Bao et al., 2021). Foliar Zn application could be a direct and operable cultivation technique for enhancing the 2-AP content and improving the taro-like flavour of pumpkin leaves during the process of agricultural production. Revealing the molecular mechanism underlying the Zn-induced changes of aroma formation and increase of 2-AP content is vital to help a more accurate and effective application of exogenous elements in improving the aroma quality of pumpkin leaves, as well as the new aromatic pumpkin cultivars through genetic engineering.

Omics are imperative in discovering how plants undergo complex biological processes as associated techniques are effective in the rapidly expanding collection of biomolecule information in plants. In this study, metabolome and transcriptome analysis were applied to pumpkin leaves at different time-points after foliar Zn treatment. The omic data was interpreted and integrated to obtain a more comprehensive understanding of involved biomolecules and their interrelationships, and their roles in 2-AP accumulation and aroma formation. The research found that Zn significantly enhanced the 2-AP concentration in pumpkin leaves. It was manifested that the Zn-induced responses in pumpkin leaves were at both metabolite level and gene level. The improved biosynthesis of 2-AP precursors, such as ornithine and proline, and associated genes contributed to the accumulation of 2-AP. Hormones and transcription factors (TFs) were involved in the regulation of 2-AP biosynthesis. Our research elucidated that Zn application could be an effective technique to improve the aromatic profile of the taro-like flavours of pumpkin leaves. The illustrated molecular mechanisms underlying Zn-induced increase of 2-AP will be of great value to the quality breeding of pumpkins associated with aroma formation.

2 Materials and methods

2.1 Plant materials

The pumpkin variety NO.44, belonging to the *Cucurbita moschata* D., was selected as the experimental material. This material is the selfed progeny of one important pumpkin germplasm that has the taro-like aroma and is widely used for pumpkin fragrance-associated breeding in China. Seeds were germinated at 30°C in the dark for two days and then cultured in Hoagland's hydroponic medium which contains all essential elements at proper levels for growth. Treatment was carefully applied to one-month-old pumpkin through evenly spraying $ZnCl_2$ (60 mL, 50 mg/L) on each leaf surface according to the previous experiments (Bao et al., 2021; Imran et al., 2022). The treatment and concentration of $ZnCl_2$ was determined according to previous studies as the specified application significantly improved the 2-AP abundance of pumpkin (not published). After foliar Zn application, fresh pumpkin leaves at top three nodes were separately collected at 24 hrs, 72 hrs, and 120 hrs. Plant samples sprayed with MilliQ water were taken as the controls. Three biological replicates were prepared for each condition, and three plants were collected for each replicate. The three experimental conditions (Zn_24 hrs, Zn_72 hrs, and Zn_120 hrs), together with the control (Ctrl), constituted the experimental materials in this study. Collected tissues were snap-frozen in liquid nitrogen tank, followed by storage inside an ultra-low temperature freezer at -80°C for the subsequent experiments.

2.2 Volatile metabolites analysis using GC-MS

Volatile compounds were analyzed *via* the headspace solid-phase microextraction (HS-SPME) combined with GC-MS analytical technology. Samples were freeze-dried in a vacuum and ground into powder. Half gram of lyophilized and powdered leave sample was weighed and carefully placed in a 20 mL vial. 3-nonanone and 2,4,6-trimethylpyridine (2 µL, 50µg/mL) were added to the vial as the dual internal standards (Fan et al., 2021; Lan et al., 2021; Li et al., 2022; Lu et al., 2022). The vial was sealed immediately and equilibrated at 70°C for five minutes. DVB/CarbonWR/PDMS Smart SPME Arrow (1.1 mm, 120/20 µm) (CTC Analytics AG) was used for the extraction of volatile metabolites. After 30 min of extraction and absorption at 70°C, SPME was desorpted and sample was analyzed on 8890-5977B GC/MS (Agilent Technologies, USA). The GC temperature program started at 35°C (2 min), then rose up to 100°C (5 min) at a rate of 5°C/min, and finally kept at 215 °C. The mass spectrometer was operated in electron impact (EI) mode with the ionization energy of 70 eV. The ion source temperature was programmed as 230°C. The full scan mode was applied with a mass scan range of 35-450 amu. The relative abundance of the compound was determined with a method related the peak areas of volatiles to that of the internal standard. The downstream comparative analysis was applied to the identified metabolites. A metabolite was considered with changed abundance if the result of statistical analysis was significant (p-value lower than 0.05).

2.3 Nonvolatile metabolites analysis using LC-MS/MS

The metabolomic profiling experiments were conducted by Metware, Wuhan. The pumpkin leave samples were processed in a vacuum freeze dryer (Scientz-100F). The freeze-dried leaves were then grounded (MM 400, Retsch). 100 mg of leave power was weighed and used for metabolite extraction. Metabolites were extracted using 70% cold methanol. Samples were vortexed for 30 sec every 30 min, after six times of vortex, the samples were kept in the extraction solution at 4°C for overnight. High speed centrifuge (10,000 g, 10 min) was followed after the overnight extraction. The supernatant was carefully transferred onto the filter membrane (0.22 µm pore size). The filtered sample was kept in vial before loading to the instrument for mass spectrometer analysis (MS, 4500 Q TRAP, Applied Biosystems, MA, United States) coupled to an ultra performance liquid chromatography (UPLC, Shim-pack UFLC SHIMADZU CBM30A system, Kyoto, Japan). Metabolites were characterized using the multiple reaction monitoring method, and then searched against the MetWare database. The downstream data analysis was carried out using Analyst 1.6.1 software. Metabolites were considered with changed abundance if the result of statistical analysis was significant (p-value lower than 0.05).

2.4 Transcriptome analysis

RNA extraction was performed using the Trizol reagent supplied by TaKaRa, Japan, according to the instructions. The RNA integrity and purity were examined through the agarose gel electrophoresis experiment. Qubit 2.0 fluorometer (Thermo Fisher Scientific, MA, USA) was used to measure the RNA concentration. The RNA quality was further confirmed using the Agilent 2100 Bioanalyzer (Agilent, CA, United States) followed by cDNA library construction. The constructed cDNA library was validated through quality control assays. Library concentration and size distribution were confirmed through Qubit dsDNA high sensitivity assay and Agilent ScreenTape assay, respectively. Kapa BioSystems qPCR method was applied to calculate molarities and normalize libraries prior to pooling and sequencing. The sequencing data was uploaded to the CNGB Sequence Archive (CNSA) of China National Genebank DataBase (CNGBdb) (accession number CNP0003866). Clean reads were aligned to the *Cucurbita moschata* reference genome (http://cucurbitgenomics.org/). Differentially expressed genes were analysed using DESeq2 package. The statistical significance testing approach was applied to identify genes that express differently (p-value less than 0.05) (Conesa et al., 2016). The functions of identified genes were annotated using the GO, KEGG and Nr databases.

2.5 Quantitative real-time PCR

The RNA and cDNA samples were well-prepared, and experiment was conducted as previously described (Deng et al., 2016). Actin was selected as the control for the normalization of gene expression levels

in pumpkin leaves (Liu et al., 2021). Primers for gene amplification (Table S8) were designed using Primer Premier 5. Experiment was conducted with the SYBR Premix Ex Taq Kit (TaKaRa, Kusatsu-Shiga, Japan) according to the protocol. The RT-qPCR reaction mixture system consisted of 1 µL cDNA, 5 µL SYBR Green PCR Master Mix (Takara), 0.3 µL of 10 µM forward primer, 0.3 µL of 10 µM reverse primer, and 3.4 µL ddH2O to a final volume 10 µL. qRT-PCR was performed on the CFX96 machine (Bio-Rad, CA, United States) under the following condition: 30 s at 95°C followed by 40 cycles (5 s at 95°C, 30 s at 60°C) and melting curve analysis at 65-95 °C. Relative expression levels of targeted genes were determined with the $2^{-\Delta\Delta CT}$ method (Deng et al., 2016). The expression level of each gene was calculated as the average of that of all three replicates from different condtions.

3 Results

3.1 Metabolite changes in pumpkin after foliar zn application

2-AP is highly volatile. According to the form that a metabolite exists, GC-MS was applied to detect the impact of Zn application on the formation of 2-AP and other aroma volatiles in pumpkin leaves. As

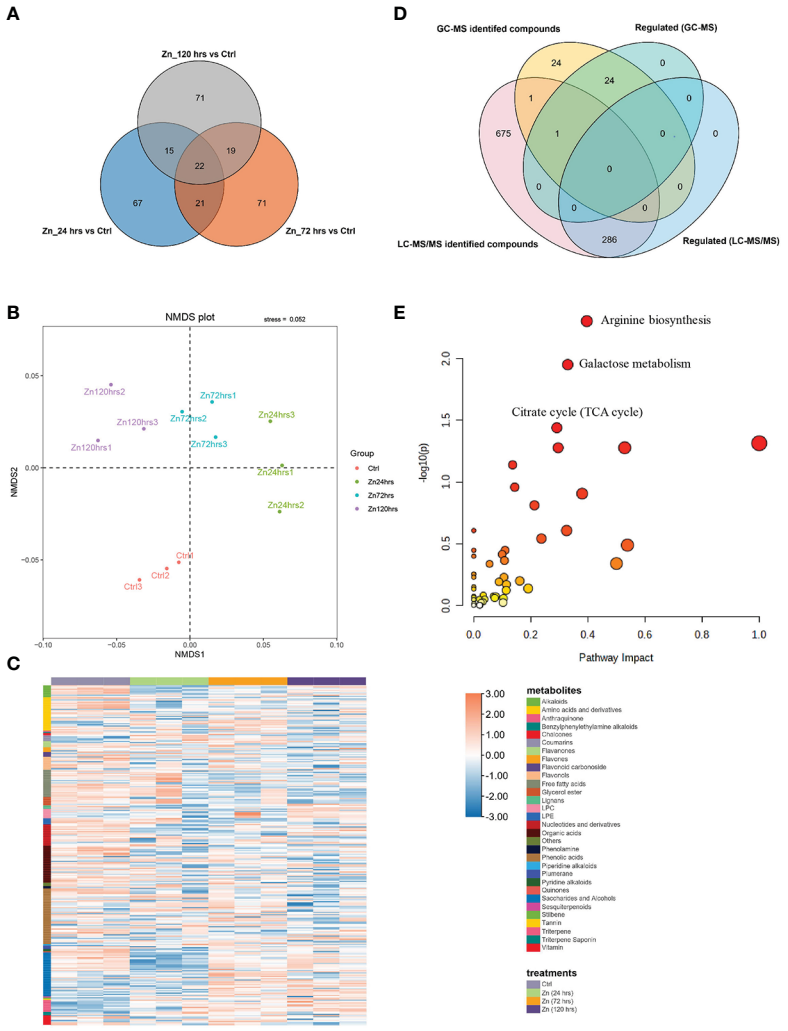


FIGURE 1
Metabolite changes in pumpkin induced by Zn application. **(A)** Venn diagram indicating the overlaps between the differentially expressed nonvolatile metabolites in the pumpkin at 24 hrs, 72 hrs, and 120 hrs after Zn treatment, respectively (p value < 0.05). **(B)** NMDS (non-metric multidimensional scaling) analysis represents the rank orders of four experimental conditions after clustering (the stress value was 0.052, which was lower than 0.1). **(C)** Heatmap generated with the Zn regulated metabolites or substances in the pumpkin. Orange and blue colors separately indicate the relative high or low abundance of metabolites in four different experimental conditions. The column and row banners represent the classifications of different treatments and component classes, respectively. **(D)** Venn diagram indicating the overlaps between the GC-MS identified volatiles and LC-MS/MS identified nonvolatiles in the pumpkin at 24 hrs, 72 hrs, and 120 hrs after Zn treatment, respectively. **(E)** The scatter plot shows enriched and matched pathways using all regulated compounds that were successfully annotated using the database in MetaboAnalyst (only the top three enriched pathways were labelled). The x and y axes are pathway impact values from the pathway topology analysis, and the corresponding p value from the pathway enrichment analysis, respectively.

expected, 2-AP was detected by GC-MS, and had significantly increased abundance at 24 hrs and 72 hrs. Including 2-AP, there were 50 volatile compounds identified from volatile emissions (Table S1), among which 25 metabolites changed their abundance after various Zn treatments. In addition, geraniol, which is an important terpene alcohol and gives some aromatic plants the rose-like pleasant fragrance, was also detected through the GC-MS approach, and increased abundance was found at 24 hrs. However, 3-hexen-1-ol, which is well-known for its strong freshly cut grass-like aroma, was decreased at all stages after Zn application. The changed abundance of different aromatic compounds determines the overall aroma profile of pumpkin.

Followingly, liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was utilized to detect metabolites in the pumpkin leaves at different time intervals after foliar Zn treatments, with the aim to integrate with the GC-MS generated data and explore the intermediates or precursors associated with the identified 2-AP and its biosynthesis. By contrast, more abundant metabolites (963 metabolites) were identified from pumpkin leave using LC-MS/MS (Table S2). Zn-induced fluctuations in metabolite generation were excavated through the downstream bioinformatic analysis. There were 286 metabolites changed in their abundance, which account for 29.7% of the identifications by LC-MS/MS (Figure 1A). Detailly, there were separately 125, 133, and 127 metabolites regulated in the pumpkin at 24 hrs, 72 hrs, and 120 hrs after the foliar application of Zn. The classification analysis through the non-metric multidimensional scaling (NMDS) method found different experimental conditions (Zn24hrs, Zn72hrs, Zn120hrs, and Ctrl) were well-clustered and separated, indicating Zn treatments affected the metabolite production (Figure 1B). The regulated metabolites were analyzed and categorized into specific groups (Figure 1C; Table S2). Phenolic acid, saccharide and alcohols, organic acids, amino acids and derivates, and free fatty acids were proved to be the largest five categories, with 47, 37, 31, 28, and 23 compounds being regulated by Zn treatment. Comparative analysis indicated that variated production of those commpounds after Zn treatment. For instance, most free fatty acids were decreased (19 out of 23 compounds), especially at 72 hrs and 120 hrs after Zn application. However, the ten regulated triterpene metabolites were all increased at different time-points after Zn treatment. There were four, four and three Zn-regulated triterpene compounds at 24 hrs, 72 hrs, and 120 hrs, respectively, with fold change ranged from 1.62 to 6.55.

Venn analysis (Figure 1D) between volatile and nonvolatile metabolites separately identified using GC-MS and LC-MS/MS indicated that there were few overlaps of identifications except for 1-decanol and benzaldehyde, demonstrating the importance of integrating different mass spectrometry methods for illustrating complex metabolic changes. Importantly, the metabolite analysis which ultilized and integrated different MS methods identified 2-AP, and 2-AP formation closely related compounds, including pyrroline, ornithine, and proline. Increased concentrations of 2-AP, ornithine, and proline were specially recognized at 24 hrs and (or) 72 hrs compared to the control, indicating the 2-AP generation and accumulation could be related to the biosynthesis and metabolism of amino acids, as the previous study reported. Compounds of which abundance were regulated by Zn treaments were further annotated, followed by pathway analysis using MetaboAnalyst (https://www.metaboanalyst.ca/). Arginine biosynthesis, galactose metabolism, and citrate cycle were the top three significantly enriched pathways (Figure 1E), and there were separately six, seven, and five compounds with regulation classified into the three pathways according to their orders mentioned above.

3.2 Gene expression dynamics in pumpkin induced by Zn

The expression of genes was investigated to explore the potential molecular mechanisms underlying the Zn induced metabolite changes, especially the regulation associated with the 2-AP production. Transcriptome analysis discovered 30,574 genes, including 953 novel genes (Table S3). Comparative analysis found 41.6% of identified genes in pumpkin leaves responded to Zn treatment. A total of 12,734 genes significantly changed their expression levels, and the most regulated genes were found at 24 hrs which was the earliest stage after the Zn application in this study (Figures 2A, B). At different time points, there were separately 8447 (4,113 increased; 4,334 decreased), 6,955 (3,503 increased; 3,452 decreased), and 5,310 (2,624 increased; 2,686 decreased) genes with changed abundance at 24 hrs (Table S4), 72 hrs (Table S5), and 120 hrs (Table S6). Additionally, the regulation of genes was impacted by Zn in a time-dependent manner. Similar yet different regulations of gene expression were discovered among distinct treatments when compared to the control condition (Figure 2C). Genes with the same regulation patterns (either increased or decreased abundance) were mostly identified between 24 hrs and 72 hrs.

3.3 Integration analysis of metabolome and transcriptome data

The metabolome and transcriptome data were integrated to illustrate the molecular changes on both metabolite and gene levels in pumpkin after Zn treatments. The enrichment analysis identified various pathways with regulations in metabolites and associated genes (Figure 3). It is evident that regulated genes were significantly enriched in pathway carbon metabolism after Zn treatments (24, 72, and 120 hrs) when compared to the normal condition. More genes with expression changes were found in pathway biosynthesis of amino acids at both 24 hrs (Figure 3A) and 72 hrs (Figure 3B). Responded genes involved in plant hormone signal transduction were especially increased in pumpkin after the Zn treatment for a relatively longer time-point (72 hrs). More remarkable fluctuations in pathway starch and sucrose metabolism and associated genes were observed at 120 hrs time-point after Zn application (Figure 3C).

Followingly, to achieve a more in-depth understanding of molecular mechanisms underlying 2-AP production and accumulation in pumpkin with taro-like flavour, further analysis was applied to integrate associated pathways, including biosynthesis of amino acid, arginine biosynthesis, TCA cycle, plant hormone signal transduction, phenylalanine metabolism, and other related pathways with direct or indirect interactions (Figures 4, 5).

10.3389/fpls.2023.1127032

3.4 Pivotal regulated genes and metabolites involved in the production of 2-AP

As mentioned above, the production of 2-AP involves several pathways, including biosynthesis of amino acids, TCA cycle, biosynthesis of arginine, and biosynthesis of arginine and proline, indicating the complex regulatory networks that involved during the process of 2-AP biosynthesis in the foliar Zn primed pumpkin. Those pathways were integrated and carefully organised to illustrate the process of 2-AP formation, and visualise the dynamic changes of associate intermediate metabolites and genes after different time intervals (Figure 4). The map comprehensively demonstrated the 2-AP formation related genes and metabolites, and revealed Zn-induced gene-to-metabolite changes. Zn significantly changed the abundance of 2-AP by regulating the associated genes. There were 11 metabolites with increased abundance at various time-points after Zn induction involved in 2-AP formation, including glutamate, N-acetyl ornithine, ornithine, and proline. In addition, 2-AP biosynthesis related vital genes, such as ornithine–oxo-acid transaminase (*OAT*), pyrroline-5-carboxylate reductase (*P5CR*), and delta-1-pyrroline-5-carboxylate synthetase (*P5CS*), were up-regulated, which results basically explained the increased abundance of 2-AP or related metabolites.

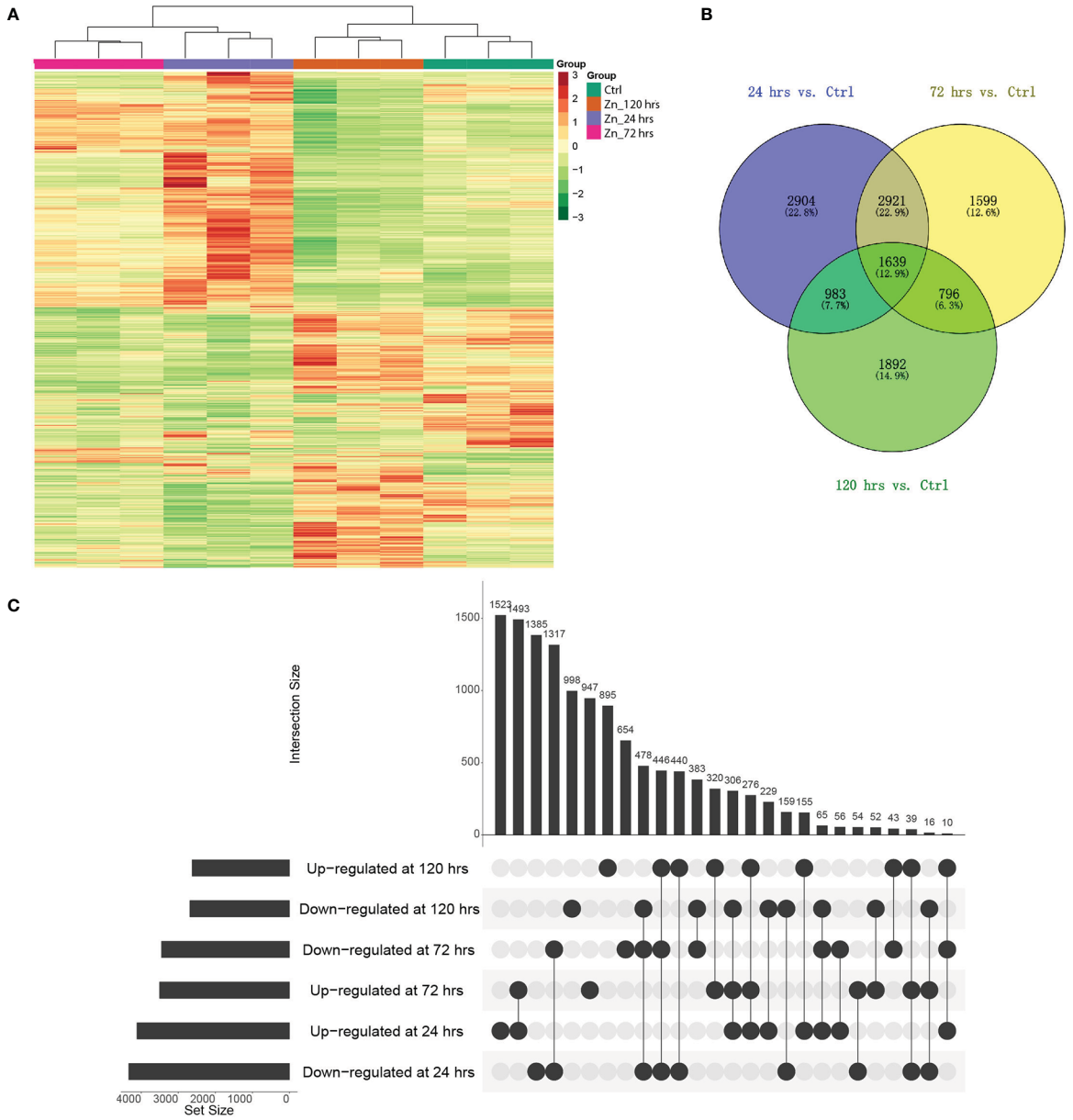


FIGURE 2
Differential gene expression in pumpkin at different time-points after Zn induction. **(A)** Heatmaps generated with differentially expressed genes after Zn treatments. Column colors indicate treatment type. Red and green colors indicate the relatively higher or lower expression levels of genes, respectively. **(B)** Venn diagram indicating the overlaps of differentially expressed genes in the pumpkin at three time points (p_value < 0.05). **(C)** UpSet visualisation of the number of regulated genes at different conditions compared to the control. Filled circles and vertical lines represent the corresponding genes being compared. The left bar graph for the UpSet plot shows the number of significantly increased or decreased gene changes in each of six comparisons upon Zn treatment.

10.3389/fpls.2023.1127032

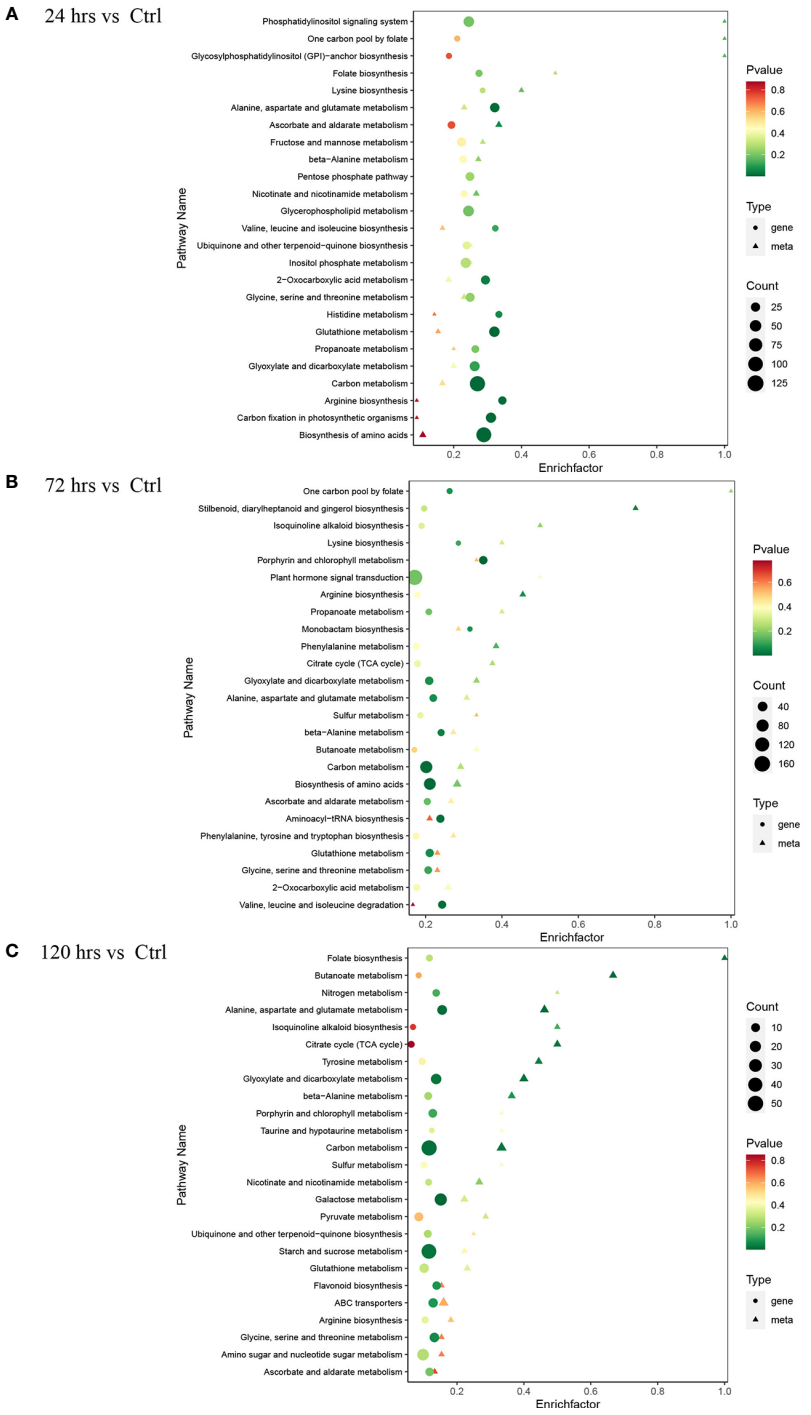


FIGURE 3
KEGG pathway enrichment analysis of differentially expressed genes and metabolites in the pumpkin at different time-points (only the top 25 significantly enriched pathways were displayed). **(A)** 24 hrs. **(B)** 72 hrs. **(C)** 120 hrs. The advanced bubble chart shows the enrichment of differentially expressed genes in various pathways. The y-axis represents pathways, and x-axis represents the enrich factors (enrich factor refers to the ratio of the number of regulated genes enriched in the pathway and the total amount of genes in the background in the same pathway). The greater rich factor indicates a higher enrichment. The size and colours (green, yellow, and red) of the bubble separately represent the amount of differentially expressed genes enriched in the pathway and its enrichment significance. The bigger bubble represents more associated genes regulated and enriched in the corresponding pathway. The smaller p-value represents better enrichment.

3.5 Plant hormone biosynthesis related molecular changes

Exogenous applications of stimulates can trigger the activation of downstream pathways *via* phytohormones homeostasis and their signaling networks. Phytohormone regulation initiates the biosynthesis of associated metabolites, including the volatile metabolites 2-AP and its precursor proline, by modulating genes involved in the corresponding biosynthetic pathway, thereby, affecting crop performance (Jogawat et al., 2021). As roughly

illustrated in Figure 5, Zn activated pathways related to the synthesis of plant hormones, including salicylic acid (SA) and abscisic acid (ABA). Our study found three SA synthesis-related pathways were impacted by Zn, including the well-known isochorismate synthase (ICS) and phenylalanine ammonia lyase (PAL) pathways in plants which both start from the chorismate. Strikingly, SA accumulated at 72 hrs after the pumpkin was sprayed with Zn, but all copies of *ICS* genes were down-regulated. In addition, all three regulated *ICSs* at 24 hrs were continuously and closely distributed on the same chromosome 4 (Figure S1). Besides, most copies of *PAL* genes were identified with decreased expression. Only one *PAL* (*CmoCh07G009550*) was up-regulated with more than three-fold changes at the initial stage (24 hrs), while the other three genes (*CmoCh10G007740*, *CmoCh20G005320* and *CmoCh03G012290*) were all down-regulated, representing the incompatible and compatible responses to Zn treatment. The third proposed pathway in SA synthesis is associated with mandelonitrile. Mandelonitrile is another SA precursor and derived from L-phenylalanine through cytochrome P450 (CYP) and mandelonitrile lyase (MDL) (Bernal-Vicente et al., 2018). Either down- or up-regulation of three *MDLs* was identified. The last step that catalyzes benzoic acid (BA) to SA is a presumed enzyme benzoic acid hydroxylase (BA2H) (Lefevere et al., 2020), but was not identified in our study. Interestingly, the regulation of methyl salicylate (MeSA) at 72 hrs was opposite to that of SA. Salicylic acid methyl transferase (SAMT) catalyzes the methylation of SA to form MeSA (Zubieta et al., 2003). The expression of two *SAMT* genes (*CmoCh03G007580* and *CmoCh07G004720*) decreased in this study at 24 hrs and 72 hrs, which result was consistent with the down-regulated MeSA. Another vital gene related to SA biosynthesis was salicylic acid-binding protein 2 (*SABP2*) (Vlot et al., 2008). The

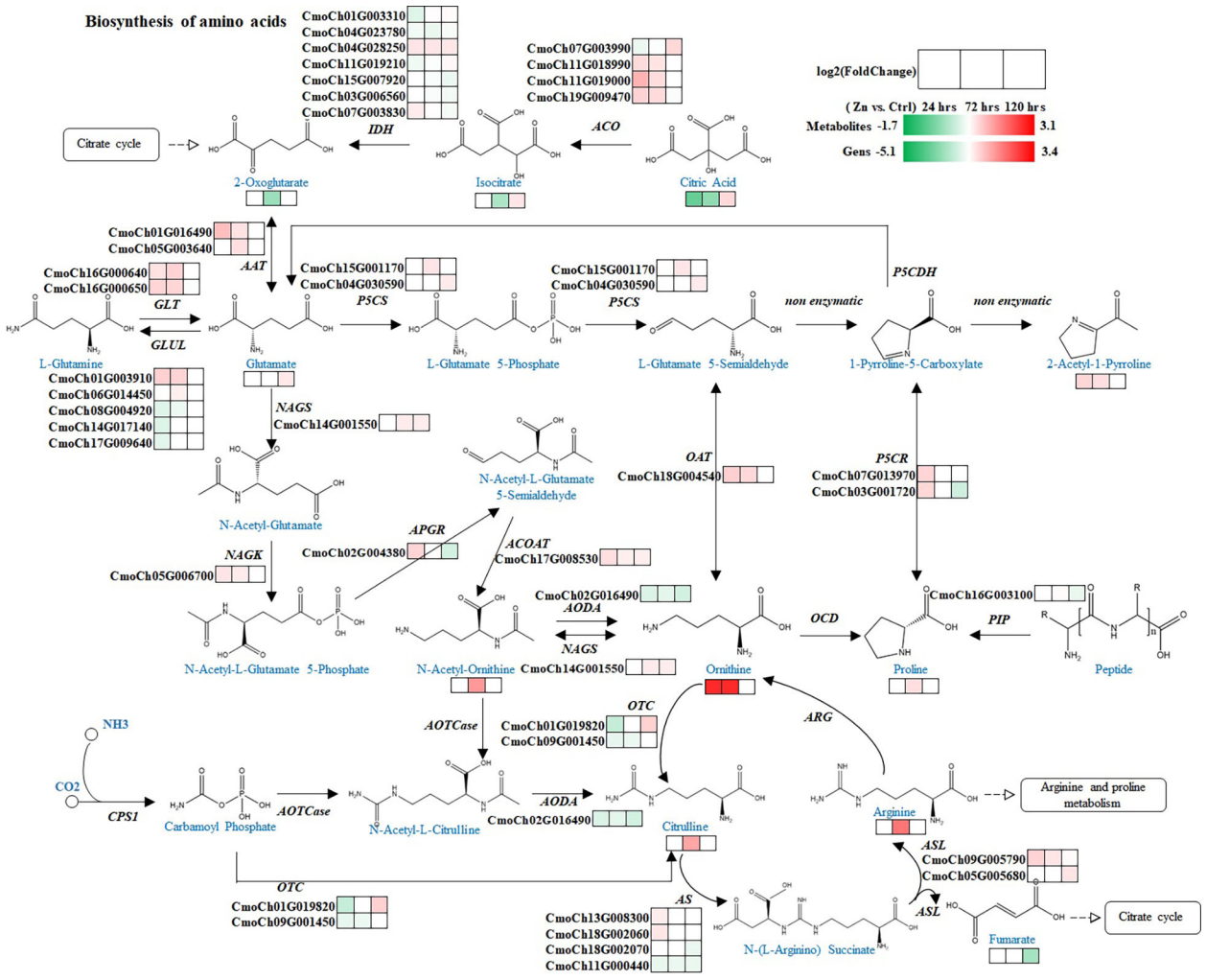


FIGURE 4
Regulation of pathways related to 2-AP biosynthesis. Blue and black colors are metabolites and genes involved in the pathway, respectively. Heatmaps illustrate the regulations of metabolites and genes at different time-points after Zn treatment. Red and green colors represent increase and decrease in abundance of metabolites or genes. Abbreviations, full names and EC numbers of enzymes: *ACO* aconitate hydratase [EC:4.2.1.3]; *IDH* isocitrate dehydrogenase [EC:1.1.1.42]; *AAT* aspartate aminotransferase [EC:2.6.1.1]; *GLT* glutamate synthase [EC:1.4.1.14]; *GLUL* glutamine synthetase [EC:6.3.1.2]; *P5CS* delta-1-pyrroline-5-carboxylate synthetase [EC:2.7.2.11, 1.2.1.41]; *NAGS* glutamate N-acetyltransferase [EC:2.3.1.35, 2.3.1.1]; *NAGK* acetylglutamate kinase [EC:2.7.2.8]; *APGR* N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]; *ACOAT* acetylornithine aminotransferase [EC:2.6.1.11]; *AODA* acetylornithine deacetylase [EC:3.5.1.16]; *OCD* ornithine cyclodeaminase [EC:4.3.1.12]; *PIP* proline iminopeptidase [EC:3.4.11.5]; OAT ornithine–oxo-acid transaminase [EC:2.6.1.13]; *P5CR* pyrroline-5-carboxylate reductase [EC:1.5.1.2]; *CPS1* carbamoyl-phosphate synthase (ammonia) [EC:6.3.4.16]; AOTCase N-acetylornithine carbamoyltransferase [EC:2.1.3.9]; *OTC* ornithine carbamoyltransferase [EC:2.1.3.3]; *AS* argininosuccinate synthase [EC:6.3.4.5]; *ASL* argininosuccinate lyase [EC:4.3.2.1]; *ARG* arginase [EC:3.5.3.1].

methyl salicylate esterase activity of SABP2 can release the active defense phytohormone SA from MeSA when active SA was required. Four *SABP2* genes were identified with differential expression in the Zn treated pumpkin. In this study, three *SABP2* displayed up-regulation, among which *CmoCh02G004360* increased at the early stage (24 hrs) after Zn application. Hence, it is supposed that the increased conversion from MeSA to SA may mostly contribute to the increased concentration of SA. Another mechanism responsible for the SA accumulation may be the decreased conversion of SA to SA 2-O-beta-D-glucoside and the glucose ester of SA (SGE) by SA-glucosyltransferase (SAGT). SAGT is the key factor that negatively regulates SA levels (Kobayashi et al., 2020). There were seven *SAGT* genes regulated by Zn. Although there were three *SAGTs* up-regulated at 24 hrs, the expression abundance of those genes was generally much lower than that of the up-regulated genes. It was assumpted that down-regulated *SAGT* may play a more predominant role in regulating the SA accumulation. Except for SA and MeSA, Zn treatment induced the accumulation of another tolerance related metabolite acetyl salicylic acid (ASA) at 120 hrs, which may be the stress tolerance mechanisms induced at the later stage.

ABA is synthesized in plants through the carotenoid pathway in which the conversion of violaxanthin to xanthoxin by 9-cis-epoxycarotenoid dioxygenase (NCED) is a rate-limiting step (Chen et al., 2020). Followingly, xanthoxin dehydrogenase (ABA2) and abscisic aldehyde oxidase (AAO3) convert xanthoxin into ABA. Four *NCED*, 10 *ABA2*, and three *AAO3* genes were found with differential expression after Zn treatment. The high number of regulated ABA2 implied there may be functional redundancy. Two of three *AAO3* were up-regulated at both 24 hrs and 72 hrs, whereas the third *AAO3* gene had increased expression only at 120 hrs.

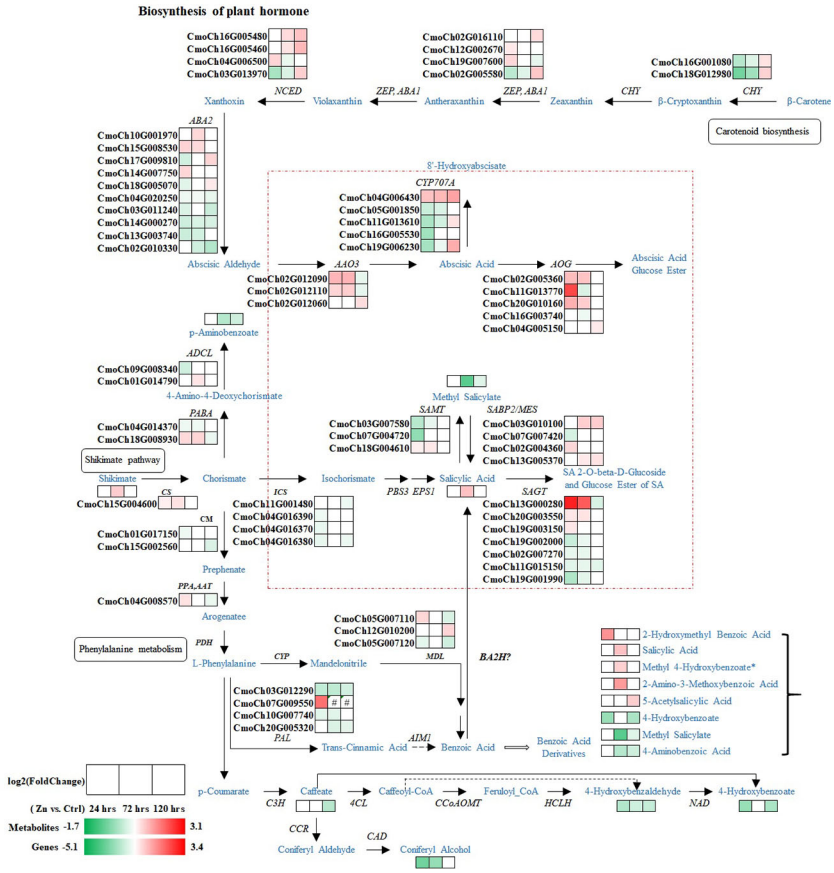


FIGURE 5
Possible biosynthesis routes for the biosynthesis of ABA and SA in the pumpkin. Blue and black colours are metabolites and genes involved in the pathway, respectively. The regulations of metabolites and genes by Zn treatments at different time-points compared to the control were demonstrated by the heatmap. Red and green colours represent the increase and decrease in abundance of metabolites or genes. The red box displays the potential core network in which relevant molecules were regulated by Zn and hence resulted in changes of ABA and SA biosynthesis in pumpkin leaves. Enzymes are abbreviated as follows: CHY carotene 3-hydroxylase [EC:1.14.15.24]; ABA1 zeaxanthin epoxidase [EC:1.14.15.21]; NCED 9-cis-epoxycarotenoid dioxygenase [EC:1.13.11.51]; ABA2 xanthoxin dehydrogenase [EC:1.1.1.288]; AAO3 abscisic aldehyde oxidase [EC:1.2.3.14]; AOG abscisate beta-glucosyl-transferase [EC:2.4.1.263]; CYP707A (+)-abscisic acid 8'-hydroxylase [EC:1.14.14.137]; ADCL 4-amino-4-deoxychorismate lyase [EC:4.1.3.38]; PABA para-aminobenzoate syn-thetase [EC:2.6.1.85]; CS chorismate synthase [EC:4.2.3.5]; ICS isochorismate synthase [EC:5.4.4.2]; PBS3 avrPphB susceptible3; EPS1 enhanced pseudomonas suscepti-bility 1; SAMT salicylic acid methyl transferase [EC: 2.1.1.274]; SABP2 salicylic acid-binding protein 2 [EC:4.1.2.10, 4.1.2.47]; SAGT SA-glucosyltransferase [EC:2.4.1.-]; CM chorismate mutase [EC:5.4.99.5]; PPA bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase [EC:2.6.1.1, 2.6.1.78, 2.6.1.79]; PDH prephenate dehydratase [EC:4.2.1.51]; CYP cytochrome P450 [EC:1.14.-.-]; MDL mandelonitrile lyase [EC:4.1.2.10]; BA2H benzoic acid 2-hydroxylase [EC1.14.13.-]. PAL phenylalanine ammonia lyase [EC:4.3.1.24]; A1M1 abnormal inflorescence meristem1; C3H p-coumarate 3-hydroxylase [EC:1.14.13.-]; 4CL 4-coumarate:CoA ligase [EC:6.2.1.12]; CCoAOMT caffeoyl-CoA O-methyltransferase [EC:2.1.1.104]; HCLH hydroxycinnamoyl-CoA hydratase-lyase; NAD benzaldehyde dehydrogenase (NAD) [EC:1.2.1.28]; CCR cinnamoyl CoA reductase [EC:1.2.1.44]; CAD cinnamyl alcohol dehydrogenase [EC:1.1.1.195].

10.3389/fpls.2023.1127032

Interestingly, the three *AAO3* genes were closely distributed on the same chromosome 3 (Figure S1). In addition, changes in ABA catabolism were discovered. Abscisic acid 8′-hydroxylase (CYP707A) catalyzes ABA to produce 8'-hydroxyabscisate (Endo et al., 2011). Five annotated *CYP707As* were identified as differentially expressed. Four of them were down-regulated at 24 hrs, while only one (*CmoCh04G006430*) was up-regulated at all time-points when compared to the control condition. Additionally, it was notable that five abscisate beta-glucosyltransferase (AOG) coded genes were regulated by Zn. In the initial of Zn treatment, three of them were regulated, and all with increased expression levels. AOG catalyzes the active ABA forming the inactive ABA conjugate ABA-glucose ester (ABA-GE), which, therefore, controls the ABA concentration and associated hormone signal in plants (Xiong et al., 2014). Although the accumulation of ABA-GE and ABA was not identified, our results indicated there may exist a tendency of increasing ABA-GE levels by promoting the generation of ABA.

3.6 Transcription factors

Responses of TFs to Zn were discovered in this research. A total of 354 TFs were identified with changed expression levels. In detail, there were 250 (65 up-regulated, 185 down-regulated), 204 (49 up-regulated,

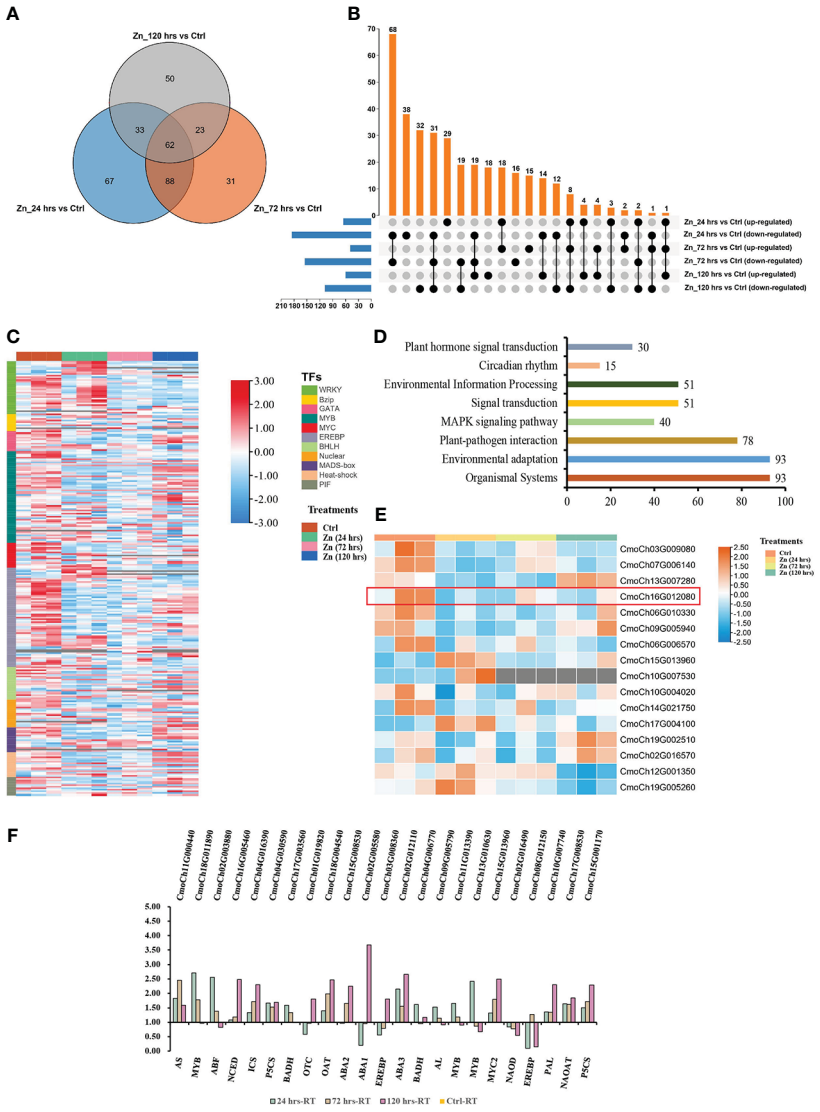


FIGURE 6
Regulated TFs in pumpkin by Zn **(A–D)**. **(A)** Venn diagram indicating the overlaps between the differentially expressed TFs in the pumpkin at 24 hrs, 72 hrs, and 120 hrs after Zn treatment, respectively. **(B)** UpSet visualisation of the number of up-regulated or down-regulated TFs in the pumpkin at three different conditions compared to the control. Filled circles and vertical lines represent the corresponding TFs (either up-regulated or down-regulated) being compared. The left bar graph for the UpSet plot shows the number of significantly increased or decreased TFs in each of the six comparisons upon Zn treatment. **(C)** Heatmap generated using the TFs with changed abundance. Red and blue colours separately indicate the relative high or low abundance in four experimental conditions. The column and row banners represent the classifications of different treatments and component classes, respectively. **(D)** Regulated TFs classified into plant hormone signal transduction and MAPK signalling pathway. **(E)** Heatmap generated with the regulated *MYC2* TFs. Orange and blue colours separately indicate the relatively high or low abundance in four experimental conditions. The column banner represents classifications of different treatments. **(F)** The expression analysis of differentially expressed genes using qRT-PCR. The x and y axises separately represent gene names and fold changes of regulated genes.

155 down-regulated), and 168 TFs regulated (60 up-regulated, 108 down-regulated) at 24 hrs, 72 hrs, and 120 hrs, respectively (Figures 6A, B). Most TFs were down-regulated in the initial (24 hrs) as a primeval response of pumpkin to the stimulates Zn. However, the regulated TFs had the trend to recover at 120 hrs (Figure 6C). *MYB*, *EREBP* and *WRKY* were the three biggest TFs groups that were regulated in Zn-treated pumpkin. KEGG enrichment analysis identified significantly enriched pathways (p-value < 0.05). Interestingly, there were separately 30 and 40 regulated TFs classified into plant hormone signal transduction and MAPK signalling pathway, indicating the involvement of these TFs in signalling after Zn induction (Figure 6D). Overlap analysis found 19 regulated TFs were involved in two mentioned pathways, including three ethylene-responsive transcription factor 1 (ERF), and 16 *MYC2* (Table S7). The TFs specially involved in plant hormone signal transduction and MAPK signalling pathway were phytochrome-interacting factor (*PIF*, 11 genes), and *WRKY* TFs (21 genes), respectively. The overlapped 16 regulated *MYC2* were further analyzed. Different regulations of various *MYC2* were identified (Figure 6E). MEGA sequence alignment and phylogenetic tree analysis (Figure S2A) demonstrated that six regulated *MYC2* in pumpkin could be homologs of *MYC2* (*AT1G32640*) in *Arabidopsis thaliana* which has been proven to involve in the negative regulation of the rate-limiting enzyme P5CS1 in the biosynthesis of proline under biotic stress (Verma et al., 2019). Expression analysis indicated that *CmoCh16G012080* was the only one with both relatively high expression, and significantly differential change after Zn induction. It was down-regulated at all three time-points, which was opposite to the regulation of *P5CS1*. NCBI CD-search prediction revealed that *CmoCh16G012080* had two domains of *AtMYC2* (Figure S2B). The prediction of protein localization sites using PSORT (https://psort.hgc.jp/form.html) elucidated that the gene was located in the nucleus of the cell, which further demonstrated the potential activities of this MYC2 on the nuclear DNA in pumpkin. The homologous sequence alignment was further confirmed using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Figure S2C). *CmoCh16G012080* and *AtMYC2* share 58.28% of identity, indicating *CmoCh16G012080* could be the potential MYC2 regulating the proline biosynthesis in pumpkins. Besides, to validate the transcriptome data, the relative expression levels of research associated genes were analyzed using qRT-PCR, and the results were proved to be reliable (Figure 6F).

4 Discussion

4.1 Zn treatments induced the gene-to-metabolite regulation in pumpkin

The integration analysis of metabolome and transcriptome comprehensively illustrated the Zn induced gene-to-metabolite regulation in pumpkin leaves. In addition, a time-dependent manner of molecule changes was observed in pumpkin after Zn priming and foliar application. Delightfully, the metabolic analysis found the 2-AP was significantly accumulated, indicating the successful induction of this compound by foliar Zn treatment, and the potential of Zn application in improving the taro-like aroma associated quality characters of pumpkin leaves. The integrated map using metabolites and genes comprehensively and firstly visualized the main pathways involved in the 2-AP formation and its correlative regulation in pumpkin leaves after Zn stimulation, and changes in abundance of related molecules, underlying there were complex regulations during the process of 2-AP production and aroma formation.

4.2 Proline and ornithine generation contributed to the accumulation of 2-AP

Illustrating the molecular mechanism of 2-AP production is important in improving the aroma quality of pumpkin leaves. 2-AP biosynthesis involves in important precursors or intermediates, and key enzymes (Bao et al., 2021). Proline is the most important precursor of 2-AP. This protein derived amino acid has been reported with multi-functions, and well-acknowledged to play an important role as an osmolyte which helps stabilize protein structure and contributes to ROS scavenging (Szabados and Savoure, 2010). This point was clarified in young plant species *Cucurbita pepo* L. and *Cucurbita moschata* Poir under salt stress. Antioxidant enzymes activities increased in the salt and proline treated plants very effectively compared to control plants under salt stress (Yoshihashi et al., 2002; Szabados and Savoure, 2010). Six superoxide dismutase (*SOD*) genes were identified as differentially expressed at different time intervals after the Zn application. Four of them were significantly up-regulated at 72 hrs, which was coincidence with the increased proline at 72 hrs.

It was reported that higher amount of proline promoted the 2-AP accumulation and improved the aroma profile of the fragrant rice (Yoshihashi et al., 2002; Bao et al., 2021). Additionally, increased 2-AP occurred in rice seedlings and callus when the intermediates or precursors were supplied, such as proline, ornithine, and glutamate. In particular, the proline increased the 2-AP concentration by more than threefold (Yoshihashi et al., 2002). Our experimental findings were consistent with these previous reports, as both increased proline and 2-AP were detected in pumpkin leaves. Thus, proline biosynthesis is vital for the 2-AP formation. Proline is synthesized predominantly from glutamate, in which enzymes P5CR and P5CS are of the utmost importance during this process (Liang et al., 2013). Both *P5CR* and *P5CS* genes had increased expression after the Zn treatment in this study. It is remarkable that ornithine was the alternative source for proline biosynthesis (Winter et al., 2015). The increased ornithine was found at both 24 hrs and 72 hrs. The yield of ornithine also starts from glutamate, followed by the acetylation, phosphorylation, reduction, and transamination (Winter et al., 2015). Vital enzymes in the pathway, including *NAGS, NAGK, APGR*, and *ACOAT* were identified, and had increased expressions, indicating those genes may regulate the formation of the downstream ornithine, citrulline, and arginine by regulating their expression patterns. Therefore, the proposed molecular mechanism related to 2-AP formation in pumpkin leaves in our study is that Zn induced the increase of 2-AP associated precursors ornithine and proline, which contribute to the 2-AP accumulation. Key genes (*P5CS*, *P5CR*, *NAGS*, *NAGK*, *APGR*, *ACOAT*, and *OAT*) were mostly up-regulated in pumpkin leaves by Zn, and therefore, positively catelyse the production of associated 2-AP precursors or intermediates. As previously described, 2-AP formation may occur in both leaves and grains of rice (Hinge et al., 2016; Pan et al., 2021). In addition, previous studies identified this component in pumpkin leaves, fruits and seeds

(Poehlmann and Schieberle, 2013; Li et al., 2019; Li et al., 2022). Potentially, findings in this study can help the better understanding of 2-AP accumulation in other tissues by keeping track of the expression of key genes and the production of precursors associated with 2-AP content, together with the synthesis and translocation of aroma volatiles across the various developmental stages.

4.3 Changes in primary metabolism played an important role during the process of aromatic compounds production

Zn, as an important micronutrient, is essential to the function of many enzymes in carbon and nitrogen metabolism, energy transfer and protein synthesis (Sung, 2020). Our study indicated that carbon metabolism was one of the most significantly enriched pathways at three time-points after the foliar Zn application. Carbon metabolism belongs to the primary metabolism, and has tight connection between the carbon metabolism and the aroma formation. Many primary metabolites that generated from primary metabolism are the direct precursors of aromatic compounds. Actually, aroma profile is the combination of hundreds of volatiles which can be generated from the primary metabolites (Pott et al., 2019).

In addition, amino acid associated pathways were also enriched by KEGG pathway analysis. Amino acid biosynthesis and metabolism are tightly connected with other biological pathways, including energy and carbohydrate metabolism, protein synthesis, carbon-nitrogen budget, stress responses, and hormone and secondary metabolism. There were 83 amino acids identified from pumpkin at 72 hrs after Zn treatment, with 14 of them being regulated (eight and six amino acids were separately down- and up-regulated). Increased levels of proline, citrulline, N-α-acetyl-L-ornithine, ornithine and arginine were particularly identified at 72 hrs, as mentioned above. Nitrogen-metabolizing pathways in plants generate ornithine, arginine, proline and polyamines through glutamate metabolism. During this process, plants produce intermediates which are vital to plant development and responses to various environmental changes (Majumdar et al., 2016). Understanding the key enzymatic reactions important for the production of primary precursors can hopefully help the transfer of metabolic flux from the central to the specialized metabolism processes, which potentially increases the formation of the final products generated from the specified biosynthetic pathway (Pott et al., 2019). Zn priming and foliar application mimicked the environmental changes. The connection between primary metabolism and aroma formation, and associated regulation are key factors for breeding crop cultivars with enhanced aroma.

4.4 Plant hormones involved in the regulation of 2-AP biosynthesis

Plant hormones are important signal molecules in sorts of signaling pathways as these pathways are directly or indirectly correlated in a wide of abiotic and biotic stress responses (Jogawat et al., 2021). The exogenous Zn, acted as the regulatory element, involves in the biosynthesis of plant hormones, and therefore, impact gene expression, could be a controversial yet interesting research direction. Zn-induced changes of genes and metabolites involved in the SA and ABA biosynthesis may affect the 2-AP formation by regulating the expression of key genes controlling the biosynthesis of specific compounds. In this study, phytohormone, as the messenger, may be involved in signal transduction systems, and induced the particular enzymes encoded genes in the proline metabolic pathway, therefore, produced associated defense compounds, including proline itself.

Kobra Maghsoudi et al. found SA elevated the expression of *P5CS*, and increased formation of proline in wheat under drought stress (Maghsoudi et al., 2018). Significantly increased SA and proline were exactly discovered at 72 hrs after Zn treatment in this study, indicating the potentially direct or indirect regulation between those two biomolecules. Three SA biosynthesis pathways were induced by Zn, but had potential different performance on SA formation. It is notable that the well-known ICS pathway in which *ICS* encodes key enzymes for SA production was negatively regulated after Zn application. All regulated *ICSs* at 24 hrs were decreased, and continuously and closely distributed on the same chromosome 4 of pumpkin, which was not identified in other species. The close distribution of genes with the same function may represent the function redundancy, or the enhanced negative regulation on ICS pathway. As reported, the predominance of different pathways related to SA biosynthesis can be different within the plants (Lefevere et al., 2020). Complex coordination among different pathways may be involved in the regulation of SA biosynthesis in pumpkin after Zn treatment in this study.

Correlation between ABA and proline was found in rice under hypoxic stress (Cao et al., 2020). ABA is believed to regulate genes relevant to proline biosynthesis, and act on upstream of the proline formation, which enhances the rice resistance and alleviates the impact of stress (Cao et al., 2020). ABA accumulated in drought-stressed plants seems to be involved in the proline synthesis by regulating the key gene *P5CS* at the transcriptional level (Bandurska et al., 2017). Although ABA accumulation was not found in this study, the associated mechanism was activated. For example, gene *AAO3* that involves in the last step of ABA biosynthesis, was discovered and up-regulated in this study. There is evidence that this gene positively regulates ABA level, and is essential for seed germination, seedling growth, grain yield, and drought tolerance in rice (Shi et al., 2021). Potentially, the biosynthesis of SA and ABA in plants can be customized so that the metabolite assembly can be redirected to specific needs.

4.5 Transcription factors acted as regulators of aromatic amino acid biosynthesis

Transcription factors involve in numerous life processes by either directly or indirectly regulating the downstream target genes. WD40, MYB, WRKY, bHLH, and bZIP families are the most common TFs among the over 60 TF families found in higher plants. These TFs were also discovered with differential expression in pumpkin leaves after Zn treatment in this study (Song et al., 2021). As described, MYB, EREBP and WRKY were the Zn-induced three largest regulated TFs groups. This finding was consistent with the results demonstrated in foliar Zn applied rice leaves where MYB and WRKY were the major differentially expressed TF families after the treatment and may play a

crucial role during 2-AP biosynthesis (Imran et al., 2022). Study found MYB TFs regulate the biosynthesis of aromatic amino acids, and the downstream secondary metabolites in rice, including the proline (Liu et al., 2015). WRKY TFs were involved in the regulation of *PAL* and *ICS* genes, which was correlated with the SA accumulation in rice (Choi et al., 2015). Notably, it was reported that *AtMYC2* in *Arabidopsis thaliana* can act as the regulatory hub within several signaling pathways. And the new role of *AtMYC2* was to regulate the *P5CS1* gene and hence proline biosynthesis (Verma et al., 2019). In this study, massive regulated *MYC2* were involved in both plant hormone signal transduction and MAPK signalling pathway. Potentially, homologs of *AtMYC2* in Zn treated pumpkin and their differential expression may be functional in the regulation of phytohormone signaling and their target genes (such as *P5CS*), therefore, regulate the proline and 2-AP synthesis. Additionally, other differentially expressed TFs in Zn treated pumpkin leaves, such as MADS-box TFs, heat shock TFs, ethylene-responsive TFs, and bHLH TFs, are also reported to be involved in the corresponding regulatory mechanisms that integrate phytohormone and redox signaling in the plant response to a number of stresses, therefore, directly or indirectly impact the 2-AP biosynthesis (Trivedi et al., 2016). Further studies are required to better understand the role of specific TFs in 2-AP biosynthesis in pumpkin leaves, especially those that directly impact the associated key genes in the pathway.

4.6 Zn induced other metabolites changes, together with 2-AP, contributed to the overall aroma profile of pumpkin

Except for increased 2-AP, the cis-3-hexen-1-ol was detected by GC-MS, but significantly down-regulated at all three time internals compared to the control. The cis-3-hexen-1-ol, as the most influential component of green aroma from plants, is a valuable resource for imparting fruit and vegetable aromas. It was reported that the hyperosmotic stress can induce the accumulation of cis-3-hexen-1-ol and up-regulation of most associated genes (such as *AHDs*) in *Camellia sinensis*, while in turn, this component could considerably improve the hyperosmotic stress tolerance through reducing the stomatal conductance and MDA, increasing ABA and proline generation (Hu et al., 2020), which pattern of regulation was different with our study. The cis-3-hexen-1-ol can be generated from 3-hexenal by the catalyzation of alcohol dehydrogenase (ADH, EC 1.1.1.1) through the alpha-linolenic acid metabolism (Jin et al., 2016). Transcriptome comparative analysis identified three *AHDs* (*CmoCh02G009540*, *CmoCh05G010180* and *CmoCh05G010190*) with decreased expression, which may result in the decreased production of cis-2-hexen-1-ol. The cis-3-hexen-1-ol with green aroma, geraniol with rose-like fragrance, 2-AP with popcorn flavour, together with other regulated aromatic compounds, contributed to the aroma profile in pumpkin leaves after Zn priming and foliar application.

5 Conclusions

Aroma of the fresh pumpkin leaves can affect the perception of consumers, and therefore the determination of their preference. The interplay of theoretical and practical have summarized that the character improvement of crops can be achieved by artificially changing the environmental conditions surrounding the plants appropriately. It is manifested that foliar application of Zn was efficient in improving the 2-AP formation in pumpkin leaves. We hypothesis that Zn modulates the 2-AP biosynthesis in a complex way at transcriptional level. The up-regulated *P5CS* after Zn treatment was supposed to contribute to the accumulated proline, and as a result, promtoed the 2-AP accumulation. It is concluded that *P5CS* is Zn-inducible, which makes it a potential gene for aroma quality improvement of pumpkin leaves. In addition, Zn may directly or indirectly interacts with plant hormones. ABA and SA, together with the transcription factors (such as MYB and MYC), were deemed to play the critical role in the elevation of 2-AP production in the metabolite profile. It is supposed that Zn application regulated the primary metabolism, which contributed to the change of aromatic components. Phytohormones may interact with Zn, and act as the signal molecule and regulate the aroma formation relevant biological processes. TFs that control specific pathways appear to be the promising tool for metabolic engineering with the aim to improve the aroma profile of pumpkins.

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: CNGB Sequence Archive (CNSA) of China National Genebank DataBase, CNP0003866.

Author contributions

JxL and XY designed and planned the experiments. JxL performed the GC-MS experiment. YQ, XY, JnL, HG, HW and XZ prepared the samples. LD, XL, GZ and JxL analyzed the data. LD and JxL wrote and reviewed the article. All authors contributed to the article and approved the submitted version.

Funding

The study was funded by the Project of Basic and Applied Basic Research of Guangdong Province [2021A1515011187], Special Fund for Scientific Innovation Strategy-Construction of High Level Academy of Agriculture Science [R2022YJ-YB1001], Key-Area Research and Development Program of Guangdong Province [2020B020220003], Agricultural Competitive Industry Discipline Team Building Project of Guangdong Academy of Agricultural Sciences [202103TD], the Science and Technology Program of Guangdong Province [2020A0505020006, 2019A050520002, 2017B030314111], the Laboratory of Lingnan Modern Agriculture Project [NZ2021008], National Modern Agricultural Technology System Construction Project [CARS-23-G-50], Construction of Modern Agricultural Industry Technology System Innovation Team

of Guangdong Province [2021KJ110], and the Science and Technology Program of Guangzhou of China [202102020801].

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.2023.1127032/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The Circos distribution of differentially expressed genes on chromosomes associated with the biosynthesis of 2-AP and hormones from and . The red and green colors separately represent the up and down regulation at 24 hrs **(A)**, 72 hrs **(B)** and 120 hrs **(C)** after Zn treatment. The size of red and green filled circles indicates the fpkm value of corresponding up- and down-regulated genes (the name on the map consists of 2AP (2-AP biosynthesis related) or Hor (hormone biosynthesis related), gene ID, and gene name), the bigger size represents higher expression level.

SUPPLEMENTARY FIGURE 2
Regulated MYC2 TFs in Zn treated pumpkin leaves. **(A)** MEGA sequence alignment and phylogenetic tree analysis using the regulated MYC2 TFs and AtMYC2. **(B)** NCBI CD-search prediction of CmoCh16G012080 and AtMYC2. **(C)** Homologous sequence alignment of CmoCh16G012080 and AtMYC2 using the online tool Clustal Omega.

References

Bandurska, H., Niedziela, J., Pietrowska-Borek, M., Nuc, K., Chadzinikolau, T., and Radzikowska, D. (2017). Regulation of proline biosynthesis and resistance to drought stress in two barley (*Hordeum vulgare* l.) genotypes of different origin. *Plant Physiol. Biochem.* 118, 427–437. doi: 10.1016/j.plaphy.2017.07.006

Bao, G., Ashraf, U., Wan, X., Zhou, Q., Li, S., Wang, C., et al. (2021). Transcriptomic analysis provides insights into foliar zinc application induced upregulation in 2-acetyl-1-pyrroline and related transcriptional regulatory mechanism in fragrant rice. *J. Agric. Food Chem.* 69, 11350–11360. doi: 10.1021/acs.jafc.1c03655

Bernal-Vicente, A., Cantabella, D., Hernandez, J. A., and Diaz-Vivancos, P. (2018). The effect of mandelonitrile, a recently described salicylic acid precursor, on peach plant response against abiotic and biotic stresses. *Plant Biol.* 20, 986–994. doi: 10.1111/plb.12894

Broadley, M. R., White, P. J., Hammond, J. P., Zelko, I., and Lux, A. (2007). Zinc in plants. *New Phytol.* 173, 677–702. doi: 10.1111/j.1469-8137.2007.01996.x

Cao, X., Wu, L., Wu, M., Zhu, C., Jin, Q., and Zhang, J. (2020). Abscisic acid mediated proline biosynthesis and antioxidant ability in roots of two different rice genotypes under hypoxic stress. *BMC Plant Biol.* 20, 198–198. doi: 10.1186/s12870-020-02414-3

Chen, K., Li, G. J., Bressan, R. A., Song, C. P., Zhu, J. K., and Zhao, Y. (2020). Abscisic acid dynamics, signaling, and functions in plants. *J. Integr. Plant Biol.* 62, 25–54. doi: 10.1111/jipb.12899

Choi, C., Hwang, S. H., Fang, I. R., Kwon, S. I., Park, S. R., Ahn, I., et al. (2015). Molecular characterization of *Oryza sativa* WRKY6, which binds to W-box-like element 1 of the *Oryza sativa* pathogenesis-related (PR) 10a promoter and confers reduced susceptibility to pathogens. *New Phytol.* 208, 846–859. doi: 10.1111/nph.13516

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., Mcpherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis (vol 17, 13, 2016). *Genome Biol.* 17 3–4, 7. doi: 10.1186/s13059-016-1047-4

Deng, L. T., Wu, Y. L., Li, J. C., Ouyang, K. X., Ding, M. M., Zhang, J. J., et al. (2016). Screening reliable reference genes for RT-qPCR analysis of gene expression in *Moringa oleifera*. *PloS One* 11, e0159458. doi: 10.1371/journal.pone.0159458

Endo, A., Kimura, M., Kawakami, N., and Nambara, E. (2011). Functional analysis of abscisic acid 8'-hydroxylase. *Methods Mol. Biol.* 773, 135–147. doi: 10.1007/978-1-61779-231-1_9

Fan, X., Jiao, X., Liu, J., Jia, M., Blanchard, C., and Zhou, Z. (2021). Characterizing the volatile compounds of different sorghum cultivars by both GC-MS and HS-GC-IMS. *Food Res. Int.* 140, 109975. doi: 10.1016/j.foodres.2020.109975

Hinge, V. R., Patil, H. B., and Nadaf, A. B. (2016). Aroma volatile analyses and 2AP characterization at various developmental stages in basmati and non-basmati scented rice (*Oryza sativa* l.) cultivars. *Rice (N Y)* 9, 38. doi: 10.1186/s12284-016-0113-6

Hu, S., Chen, Q., Guo, F., Wang, M., Zhao, H., Wang, Y., et al. (2020). (Z)-3-Hexen-1-ol accumulation enhances hyperosmotic stress tolerance in *Camellia sinensis*. *Plant Mol. Biol.* 103, 287–302. doi: 10.1007/s11103-020-00992-2

Imran, M., Liu, Y., Shafiq, S., Abbas, F., Ilahi, S., Rehman, N., et al. (2022). Transcriptional cascades in the regulation of 2-AP biosynthesis under zn supply in fragrant rice. *Plant Physiol.* 174, e13721. doi: 10.1111/ppl.13721

Jin, Y., Zhang, C., Liu, W., Tang, Y., Qi, H., Chen, H., et al. (2016). The alcohol dehydrogenase gene family in melon (*Cucumis melo* l.): bioinformatic analysis and expression patterns. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00670

Jogawat, A., Yadav, B., Chhaya,, Lakra, N., Singh, A. K., and Narayan, O. P. (2021). Crosstalk between phytohormones and secondary metabolites in the drought stress tolerance of crop plants: A review. *Plant Physiol.* 172, 1106–1132. doi: 10.1111/ppl.13328

Kobayashi, Y., Fukuzawa, N., Hyodo, A., Kim, H., Mashiyama, S., Ogihara, T., et al. (2020). Role of salicylic acid glucosyltransferase in balancing growth and defence for optimum plant fitness. *Mol. Plant Pathol.* 21, 429–442. doi: 10.1111/mpp.12906

Lan, T., Gao, C., Yuan, Q., Wang, J., Zhang, H., Sun, X., et al. (2021). Analysis of the aroma chemical composition of commonly planted kiwifruit cultivars in China. *Foods* 10 4–7. doi: 10.3390/foods10071645

Lefevere, H., Bauters, L., and Gheysen, G. (2020). Salicylic acid biosynthesis in plants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00338

Liang, X., Zhang, L., Natarajan, S. K., and Becker, D. F. (2013). Proline mechanisms of stress survival. *Antioxid Redox Signal* 19, 998–1011. doi: 10.1089/ars.2012.5074

Li, J., Miao, A., Zhao, G., Liu, X., Wu, H., Luo, J., et al. (2022). Assessment of the 'taro-like' aroma of pumpkin fruit (*Cucurbita moschata* d.) *via* e-nose, GC–MS and GC-O analysis. *Food Chemistry: X* 15, 100435. doi: 10.1016/j.fochx.2022.100435

Liu, J., Osbourn, A., and Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012

Liu, W., Zhang, R., Xiang, C., Zhang, R., Wang, Q., Wang, T., et al. (2021). Transcriptomic and physiological analysis reveal that alpha-linolenic acid biosynthesis responds to early chilling tolerance in pumpkin rootstock varieties. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.669565

Li, J., Zhong, Y., Luo, J., He, X., Gong, H., Wu, H., et al. (2019). Detection of volatile flavor compounds in leaf of xiangyu pumpkin using head space solid phase microextraction-gas chromatography-mass spectrometry. *J. Zhejiang Univ. (Agriculture Life Sciences)* 45, 175–180. doi: 10.3785/j.issn.1008-9209.2017.11.231

Lu, K., Liu, L., Xu, Z., and Xie, W. (2022). The analysis of volatile compounds through flavoromics and machine learning to identify the origin of traditional Chinese fermented shrimp paste from different regions. *LWT* 171 2–3. doi: 10.1016/j.lwt.2022.114096

Maghsoudi, K., Emam, Y., Niazi, A., Pessarakli, M., and Arvin, M. J. (2018). P5CS expression level and proline accumulation in the sensitive and tolerant wheat cultivars under control and drought stress conditions in the presence/absence of silicon and salicylic acid. *J. Plant Interact.* 13, 461–471. doi: 10.1080/17429145.2018.1506516

Majumdar, R., Barchi, B., Turlapati, S. A., Gagne, M., Minocha, R., Long, S., et al. (2016). Glutamate, ornithine, arginine, proline, and polyamine metabolic interactions: the pathway is regulated at the post-transcriptional level. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00078

Mashiane, P., Mashitoa, F. M., Slabbert, R. M., and Sivakumar, D. (2021). Impact of household cooking techniques on colour, antioxidant and sensory properties of African pumpkin and pumpkin leaves. *Int. J. Gastronomy Food Sci.* 23 1. doi: 10.1016/j.ijgfs.2021.100307

Pan, Y. Y., Chen, Y. B., Wang, C., H, Li, Huang., D., Zhou., D., et al. (2021). Metabolism of γ-aminobutyrate and 2-acetyl-1-pyrroline analyses at various grain developmental stages in rice. *Chin. J. Rice Sci.*, 35(2) 121–129. doi: 10.16819/j.1001-7216.2021.0805

Poehlmann, S., and Schieberle, P. (2013). Characterization of the aroma signature of styrian pumpkin seed oil (*Cucurbita pepo* subsp. pepo var. styriaca) by molecular sensory science. *J. Agric. Food Chem.* 61, 2933–2942. doi: 10.1021/jf400314j

Pott, D. M., Osorio, S., and Vallarino, J. G. (2019). From central to specialized metabolism: an overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00835

Shi, X., Tian, Q., Deng, P., Zhang, W., and Jing, W. (2021). The rice aldehyde oxidase *OsAO3* gene regulates plant growth, grain yield, and drought tolerance by participating in ABA biosynthesis. *Biochem. Biophys. Res. Commun.* 548, 189–195. doi: 10.1016/j.bbrc.2021.02.047

Song, C.-Z., Liu, M.-Y., Meng, J.-F., Shi, P.-B., Zhang, Z.-W., and Xi, Z.-M. (2016). Influence of foliage-sprayed zinc sulfate on grape quality and wine aroma characteristics of merlot. *Eur. Food Res. Technol.* 242, 609–623. doi: 10.1007/s00217-015-2570-3

Song, M., Wang, H., Wang, Z., Huang, H., Chen, S., and Ma, H. (2021). Genome-wide characterization and analysis of bHLH transcription factors related to anthocyanin biosynthesis in fig (*Ficus carica* l.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.730692

Sung, J. (2020). Metabolic profiling reveals altered carbon and nitrogen metabolism in cabbage (*Brassica rapa* l.) plants by the deficiency of micronutrients. *Korean J. Soil Sci. Fertilizer* 53, 355–365. doi: 10.7745/KJSSF.2020.53.3.355

Szabados, L., and Savoure, A. (2010). Proline: A multifunctional amino acid. *Trends Plant Sci.* 15, 89–97. doi: 10.1016/j.tplants.2009.11.009

Trivedi, D. K., Gill, S. S., and Tuteja, N. (2016). "Abscisic acid (ABA): biosynthesis, regulation, and role in abiotic stress tolerance," in *Abiotic stress response in plants*. Eds. N. TUTEJA and S. S. GILL (Germany: Wiley Wiley-VCH Verlag GmbH & Co. Weinheim).

Verma, D., Jalmi, S. K., Bhagat, P. K., Verma, N., and Sinha, A. A.-O. (2019). A bHLH transcription factor, MYC2, imparts salt intolerance by regulating proline biosynthesis in *Arabidopsis*. *FEBS J.* 287(12) doi: 10.1111/febs.15157

Vlot, A. C., Liu, P. P., Cameron, R. K., Park, S. W., Yang, Y., Kumar, D., et al. (2008). Identification of likely orthologs of tobacco salicylic acid-binding protein 2 and their role in systemic acquired resistance in *Arabidopsis thaliana*. *Plant J.* 56, 445–456. doi: 10.1111/j.1365-313X.2008.03618.x

Winter, G., Todd, C. D., Trovato, M., Forlani, G., and Funck, D. (2015). Physiological implications of arginine metabolism in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00534

Xiong, D. M., Liu, Z., Chen, H., Xue, J. T., Yang, Y., Chen, C., et al. (2014). Profiling the dynamics of abscisic acid and ABA-glucose ester after using the glucosyltransferase UGT71C5 to mediate abscisic acid homeostasis in *Arabidopsis thaliana* by HPLC-ESI-MS/MS. *J. Pharm. Anal.* 4, 190–196. doi: 10.1016/j.jpha.2014.01.004

Yoshihashi, T., Huong, N. T., and Inatomi, H. (2002). Precursors of 2-acetyl-1-pyrroline, a potent flavor compound of an aromatic rice variety. *J. Agric. Food Chem.* 50, 2001–2004. doi: 10.1021/jf011268s

Zeng, L., Watanabe, N., and Yang, Z. (2019). Understanding the biosyntheses and stress response mechanisms of aroma compounds in tea (*Camellia sinensis*) to safely and effectively improve tea aroma. *Crit. Rev. Food Sci. Nutr.* 59, 2321–2334. doi: 10.1080/10408398.2018.1506907

Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J. P. (2003). Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell* 15, 1704–1716. doi: 10.1105/tpc.014548

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fpls.2023.1090009

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Wenge Liu,
Zhengzhou Fruit Research Institute (CAAS), China
Shengping Zhang,
Insititute of Vegetables and Flowers (CAAS), China

*CORRESPONDENCE
Zhaoyi Dai
daizhaoyi01@163.com

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

RECEIVED 04 November 2022
ACCEPTED 09 January 2023
PUBLISHED 07 February 2023

CITATION
Yi L, Zhou W, Zhang Y, Chen Z, Wu N, Wang Y and Dai Z (2023) Genetic mapping of a single nuclear locus determines the white flesh color in watermelon (*Citrullus lanatus* L.).
Front. Plant Sci. 14:1090009.
doi: 10.3389/fpls.2023.1090009

Genetic mapping of a single nuclear locus determines the white flesh color in watermelon (*Citrullus lanatus* L.)

Licong Yi[1,2], Wei Zhou[1,2], Yi Zhang[3], Zibiao Chen[1,4], Na Wu[1], Yunqiang Wang[1,2] and Zhaoyi Dai[1,2]*

[1]Hubei Key Laboratory of Vegetable Germplasm Enhancement and Genetic Improvement, Industrial Crops Institute, Hubei Academy of Agricultural Science, Wuhan, China, [2]Key Laboratory of Ecological Cultivation on Alpine Vegetables (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Wuhan, China, [3]Hubei Key Laboratory of Optical Information and Pattern Recognition, Wuhan Institute of Technology, Wuhan, China, [4]College of Horticulture and Gardening, Yangtze University, Jingzhou, China

Introduction: Flesh color is an important trait in watermelon (Citrullus lanatus L.). Several flesh color genes have been identified in watermelon; however, the inheritance of and the molecular basis underlying the white flesh trait remain largely unknown.

Methods: In this study, segregation populations were constructed by crossing the canary yellow flesh line HSH-F with the white flesh line Sanbai to fine-map the white flesh gene in watermelon.

Results: Genetic analysis indicated that the white flesh trait is controlled by a single recessive locus, termed Clwf2. Map-based cloning delimited the Clwf2 locus to a 132.3-kb region on chromosome 6. The candidate region contains 13 putative genes, and four of them—Cla97C06G121860, Cla97C06G121880, Cla97C06G121890, and Cla97C06G121900—were significantly downregulated in the white flesh compared to the canary yellow flesh watermelon fruits. The Cla97C06G121890 gene, which encodes a tetratricopeptide repeat protein, showed almost no expression in the white flesh fruit before maturity, whereas it had a very high expression in the canary yellow flesh fruit at 18 days after pollination. Transmission electron microscopy revealed rounded and regularly shaped chromoplasts in both the canary yellow and white flesh fruits. Further quantitative real-time PCR analysis showed that the expression levels of several key plastid division genes and almost the entire carotenoid biosynthesis pathway genes were downregulated in the white flesh compared to the canary yellow flesh fruits.

Discussion: This study suggests that the proliferation inhibition of chromoplasts and downregulation of the CBP genes block the accumulation of carotenoids in watermelon and lead to white flesh. These findings advance and extend the understanding of the molecular mechanisms underlying white flesh trait formation and carotenoid biosynthesis in watermelon.

KEYWORDS

watermelon, genetic mapping, flesh color, carotenoids, chromoplast

Introduction

Watermelon (*Citrullus lanatus* L., $2n = 2x = 22$) is one of the most popular summer fruits grown all over the world, and it provides rich nutrients in the human diet, including carotenoids, citrulline, and vitamins. Carotenoids are important secondary metabolites that yield the red, orange, and yellow colors in many fruits, flowers, and vegetables (Nisar et al., 2015). In watermelon, the red, orange, and yellow flesh colors are abundant in lycopene, β-carotene, and xanthophyll, respectively (Zhao et al., 2013; Lv et al., 2015). Investigating inheritance of the flesh color and the molecular mechanisms underlying carotenoid accumulation is of significant importance for watermelon breeding.

There are a variety of flesh colors in watermelon, including white, pale yellow, salmon yellow, canary yellow, orange, pink, red, and scarlet red. Researchers have been investigating the inheritance of flesh color in watermelon since 1937 (Porter, 1937), and several flesh color loci have been identified (Poole, 1944; Henderson, 1989; Henderson et al., 1998; Gusmini and Wehner, 2006; Bang et al., 2007; Bang et al., 2010). It has been reported that the *C* and *I* loci determine the canary yellow and red flesh colors in watermelon (Poole, 1944). A study by Henderson et al. (1998) showed that the canary yellow flesh (*C*) is dominant over the red flesh (*c*), but can be inhibited by the homozygous recessive *i* locus, which will result in a red flesh regardless of the *C* alleles present (Henderson et al., 1998). Moreover, Shimotsuma (1963) suggested that the yellow flesh (*B*) and white flesh (*Wf*) loci, with epistatic interaction, control the yellow, red, and white flesh colors in watermelon. Yellow flesh (*B*) is dominant over red flesh (*b*), and *Wf* is epistatic to the yellow and red flesh colors. However, whether *C* and *B* belong to the same locus is still unknown. The red, orange, and salmon yellow flesh colors are controlled by the *Y* locus, which has three alleles, in which *Y* (red flesh) is dominant over both *y°* (orange flesh) and *y* (salmon yellow flesh), and *y°* (orange flesh) is dominant over *y* (salmon yellow flesh) (Henderson, 1989).

With the development of high-throughput sequencing technology and the publication of high-quality watermelon reference genome sequence (Guo et al., 2012; Guo et al., 2019), great breakthroughs have been made in the mapping and cloning of the flesh color genes in watermelon. Branham et al. (2017) mapped a major codominant quantitative trait locus (QTL) (i.e., *QFC.1*) that is associated with the accumulation of β-carotene on watermelon chromosome 1 based on linkage analysis. Zhang et al. (2017) revealed that the expression of the phosphorus transporter *ClPHT4;2* is essential for pigment accumulation in watermelon. Guo et al. (2019) identified five flesh color loci in watermelon using genome-wide association study (GWAS), in which the *FC4.1* locus on chromosome 4 contains a lycopene β-cyclase (*LCYB*) gene. Mutation in the *LCYB* gene has been reported to lead to flesh color changes from lemon yellow to red (Bang et al., 2007). Li et al. (2020) identified a locus (i.e., Y^{scr}) on chromosome 6 that determines the scarlet red flesh color trait in watermelon. Diao et al. (2021) located a locus for the canary yellow flesh color on chromosome 6. In addition, Wang et al. (2021) mapped a locus associated with the pale yellow and white flesh color variations on chromosome 1. The expression level of *ClPSY1* was positively correlated with lycopene accumulation in watermelon (Guo et al., 2019), and a nucleotide variation in the *ClPSY1* gene resulted in a golden flesh color (Liu et al., 2022). Furthermore, comparative transcriptome studies showed that low expression of the lycopene cyclase genes (*LCYB/LCYE*) during fruit development is a key factor leading to lycopene accumulation in watermelon (Grassi et al., 2013; Lv et al., 2015).

The carotenoid biosynthesis pathway (CBP) in plants has been extensively studied, and the major genes and enzymes in this pathway have been well characterized (Cazzonelli and Pogson, 2010; Nisar et al., 2015; Yuan et al., 2015; Sun et al., 2018). The C_{20} glyceraldehyde 3-phosphate (GGPP) molecule, which is produced from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) of the methylerythritol 4-phosphate (MEP) pathway, is the immediate precursor of carotenoids (Fraser and Bramley, 2004). *PSY* catalyzes the condensation of two GGPP molecules to form C_{40} carotenoid phytoene. Phytoene is desaturated and isomerized to produce lycopene under catalysis of phytoene desaturase (PDS), ζ-carotene dehydrogenase (ZDS), ζ-carotene isomerase (Z-ISO), and ζ-carotene desaturase (CRTISO) (Nisar et al., 2015). Lycopene is cyclized by *LCYB* and *LCYE* to produce β-carotene and α-carotene, respectively. These carotenes are further converted to xanthophylls through a series of hydroxylation reactions. In addition to the CBP genes, several transcriptional regulators, such as MYB, WRKY, MADS, bHLH, B-box, ERF, and NAC, that control the accumulation of carotenoids by directly interacting with the CBP genes to repress or activate their expression have been identified (Welsch et al., 2007; Ma et al., 2014; Larkin et al., 2016; Sagawa et al., 2016; Fu et al., 2017; Lu et al., 2018; Ampomah-Dwamena et al., 2019; Meng et al., 2019; Xiong et al., 2019; Zhou et al., 2019; Duan et al., 2022). For example, WHITE PETAL1 (WP1), a R2R3-MYB protein, plays a crucial role in regulating floral carotenoid pigmentation in *Medicago* by directly regulating the expression of *MtLYCe* and *MtLYCb* (Meng et al., 2019). In kiwifruit, *TFMYB7* controls pigment accumulation during fruit ripening by activating the promoter of the *AdLCYb* gene (Ampomah-Dwamena et al., 2019). In tomato, *SlNAC1* interacts with the *SlPSY1* gene to alter the carotenoid pathway flux (Ma et al., 2014).

Colored flesh watermelon cultivars evolved from white flesh wild watermelon, and human selection of CBP genes, particularly the *PSY* and *LCYB* genes, have played a key role in flesh coloration in watermelon (Guo et al., 2019). In the present study, we report the mapping and candidate gene prediction of the *Clwf2* locus that determines the white flesh trait in watermelon. Genes related to plastid proliferation and carotenoid biosynthesis were significantly downregulated in white flesh watermelon fruits. The findings of this study provide new insights into flesh color formation and promote understanding of carotenoid accumulation in watermelon.

Materials and methods

Plant materials and genetic populations

The watermelon germplasms HSH-F, Sanbai, and CG149 are preserved by the Industrial Crops Institute of Hubei Academy of Agricultural Sciences, Wuhan. The flesh colors of these germplasms are canary yellow, white, and orange, respectively.

For genetic inheritance analysis and genetic mapping of the white flesh trait, the canary yellow flesh inbred line HSH-F (P_1) was crossed with the white flesh line Sanbai (P_2) to produce F_1 hybrids. Subsequently, the F_1 plants were backcrossed with HSH-F and Sanbai to produce the BC_1P_1 and BC_1P_2 populations and were self-pollinated to generate the F_2 segregation population. F_2 individuals were further self-pollinated and individually harvested to generate F_3 families. The F_2 segregation population was used for genetic mapping of the white flesh gene. To examine whether the white and canary yellow flesh traits are controlled by the same locus, the orange flesh line CG149 (P_3) was crossed with HSH-F and Sanbai to produce the $F_1(P_3\times P_1)$ and $F_1(P_3\times P_2)$ hybrids, respectively. The plant materials were creeping-cultured in a greenhouse under natural light conditions at the vegetable test base of Hubei Academy of Agricultural Sciences, Wuhan (30°36′ N, 104°18′ E). The greenhouse was maintained at daily temperatures between 20°C and 40°C and relative humidity between 50% and 85%. Female flowers were pollinated from 7:00 to 9:00 a.m. The flesh colors of the watermelon fruits were visually recorded at 34 days after pollination (DAP).

Genome resequencing and variant calling

Genomic DNA was extracted from the leaves of HSH-F and Sanbai using the cetyltrimethyl-ammonium bromide (CTAB) method (Porebski et al., 1997). The integrity of genomic DNA was examined using 1% agarose gel, purity was assessed with a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the concentrations were determined using Qubit 3.0 (Thermo Fisher Scientific). A DNA library was constructed using the NEB_Next Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Hitchin, UK) according to the manufacturer's instructions and then sequenced using the HiSeq 4000 (Illumina, San Diego, CA, USA) platform with a paired-end run (2 × 150 bp). After removing the adaptors and low-quality reads, the clean reads were mapped to the watermelon reference genome 97103 (v2) (http://cucurbitgenomics.org/organism/21) using the Burrows–Wheeler Aligner (BWA, v0.7.15-r1140) software. The Genome Analysis ToolKit (GATK, v3.7) was used for single nucleotide polymorphism (SNP) and insertion–deletion (InDel) calling.

Genetic analysis of the flesh color trait

For genetic inheritance analysis, the F_2 (n = 137), BC_1P_1 (n = 39), and BC_1P_2 (n = 41) segregation populations derived from the cross between HSH-F and Sanbai were grown in the spring of 2019, and the flesh colors were recorded individually at 34 DAP. To validate the genetic inheritance of flesh color, 51 F_3 families were grown in the fall of 2019 for phenotyping and segregation analysis. For each F_3 family, the flesh colors of 15 individuals were recorded. The deviation from the expected segregation ratio of each population was analyzed using the chi-square test.

Marker development and genetic mapping

For preliminary mapping of the white flesh locus, an SNP panel containing 390 SNPs distributed evenly across 11 watermelon chromosomes was developed based on published watermelon resequencing data (ftp://cucurbitgenomics.org/pub/cucurbit/reseq/watermelon/v2/). The primer sequences and genomic positions of the 390 SNPs are listed in Supplementary Table S1. The F_2 population with 137 individuals and the 51 F_3 families derived from the cross between HSH-F and Sanbai were firstly genotyped using the SNP panel. Based on the phenotypes of the F_2 and F_3 populations, QTL analysis was performed by composite interval mapping (CIM) using WinQTLCart2.5 software (Li et al., 2007). The logarithm of the odds (LOD) threshold was set at 2.5. To narrow the QTL region (Chr06:15.92–24.50 Mb), six InDel markers (Supplementary Table S1) with an average density of approximately 1.2 Mb were designed and used to genotype the 137 F_2 individuals.

According to the preliminary mapping results, two flanking InDel markers were used to screen recombinant plants from the 1,054 F_2 individuals at the seedling stage in the spring of 2020. These recombinants were further transplanted in the greenhouse for phenotyping. Seven SNP markers (Supplementary Table S1) with an average density of about 100 kb were designed and used to genotype the recombinant plants in order to fine-map the white flesh locus.

Functional annotation of candidate genes

Functional annotation of the genes in the fine-mapped region was based on watermelon reference genome 97103 (v2) (ftp://cucurbitgenomics.org/pub/cucurbit/reseq/watermelon/v2/).

Quantitative real-time PCR analysis

Flesh samples from the HSH-F and Sanbai lines were collected from the center of three parental fruits at 10, 18, 26, and 34 DAP. The flesh tissues were immediately frozen in liquid nitrogen and stored at −80°C for RNA extraction. Total RNA was extracted using an RNAprep Pure Plant Kit (BioTeKe, Beijing, China) according to the manufacturer's instructions. The concentrations of RNA were determined using NanoDrop spectrometry. Reverse transcription was performed using a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The expression levels of the candidate genes were determined by quantitative real-time PCR (qRT-PCR) assays using the MonAmp Fast SYBR Green qPCR Mixture (Monad, Wuhan, China) with the QuantStudio 5 system (ABI, Foster City, CA, USA). The *ClACT* (*Cla97C02G026960*) gene was used as an internal reference. The relative expression levels were calculated using the $2^{-(\Delta\Delta Ct)}$ method (Livak and Schmittgen, 2001). Each test had three biological and technical replicates. The genes and primers used for qRT-PCR analysis are listed in Supplementary Table S1. Student's t-test was used to assess statistical differences, and a p-value of 0.05 was used to determine the significance of the differences between means.

Detection of carotenoids

Flesh samples from the HSH-F, Sanbai, and CG149 lines were collected from the center of three parental fruits at 34 DAP. The flesh tissues were cut into small pieces, immediately frozen in liquid nitrogen, and then stored at −80°C. Subsequently, the samples were freeze-dried, ground into powder (30 Hz, 1.5 min), and stored at −80° C until use. About 50 mg powder was weighed and extracted with 0.5 ml mixed solution of *n*-hexane/acetone/ethanol (1:1:1, *v*/*v*/*v*) and the extract vortexed for 20 min at room temperature. Supernatants were then collected after centrifugation at 12,000 rpm for 5 min at 4° C. The residue was re-extracted by repeating the above steps under the same conditions, evaporated to dryness, and reconstituted in a mixed solution of methyl alcohol/methyl *tert*-butyl ether (1:1, *v*/*v*). The solution was filtered using a 0.22-μm membrane filter for further liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis. The carotenoid content was determined using MetWare (http://www.metware.cn/) based on the AB SciexQTRAP 6500 LC-MS/MS platform. Three biological repeats were performed for each line.

Transmission electron microscopy of chromoplasts

Transmission electron microscopy (TEM) was used to observe the morphology and ultrastructure of the chromoplasts in the fruits of HSH-F and Sanbai at 34 DAP. Flesh samples from the central areas were cut into 2 mm × 2 mm pieces and immediately fixed with 2.5% glutaraldehyde. After embedding in Spurr's resin, ultrathin sections of the samples were obtained with a Leica EM UC6 ultramicrotome (Leica Microsystems, Wetzlar, Germany). The sections were examined and photographed with an H-7650 transmission electron microscope (Hitachi, Tokyo, Japan) on the electron microscopy platform of Huazhong Agricultural University.

Results

White flesh accumulates trace amounts of carotenoids

The flesh color of the two parents, HSH-F and Sanbai, was observed at 10, 18, 26, and 34 DAP (Figure 1A). Pigment accumulation could be visually observed in the canary yellow flesh parent at 18 DAP, and a distinguishable flesh color was observed in the canary yellow and white flesh parents at 26 DAP. The LC-MS/MS-based determination of the carotenoid content in the fruits of the two parents at 34 DAP showed that the canary yellow flesh contains abundant xanthophyll, whereas the white flesh contains trace amounts of carotenoids (Figure 1B and Supplementary Table S2). In the canary yellow flesh, 30 types of carotenoids were identified, and most of these were xanthophyll. The five most abundant carotenoids in the canary yellow flesh were violaxanthin-myristate-laurate (7.68 ± 1.80 μg/g), violaxanthin palmitate (6.41 ± 2.50 μg/g), lutein (6.35 ± 1.49 μg/g), violaxanthin myristate (4.77 ± 1.32 μg/g), and violaxanthin-myristate-caprate (2.96 ± 1.24 μg/g) (Supplementary Table S2). In white flesh, 19 types of carotenoids were identified, but the carotenoid content was extremely low (Supplementary Table S2).

The white flesh trait is controlled by a single nuclear locus and is epistatic to the canary yellow flesh trait

Previous studies have implied an interaction effect between the white flesh and yellow flesh loci (Henderson, 1989; Henderson et al., 1998). In this study, three parents with canary yellow (HSH-F, P_1), white (Sanbai, P_2), and orange (CG149, P_3) flesh colors (Figure 2A) were used for genetic analysis of the flesh color in watermelon. Determination of the carotenoid content revealed that carotenes were highly abundant in the orange flesh parent CG149, which lacked xanthophyll (Figure 1B and Supplementary Table S2). The three most abundant carotenes in CG149 were (E/Z)-phytoene (42.94 ± 1.46 μg/g), lycopene (41.65 ± 3.74 μg/g), and β-carotene (37.00 ± 1.40 μg/g) (Supplementary Table S2).

The canary yellow flesh trait was observed in the fruits of the F_1 hybrids produced from crossing the canary yellow with the white flesh parents (HSH-F × Sanbai) (Figure 2B). The F_2 segregation population derived from the cross between HSH-F and Sanbai showed pure yellow, asymmetrical yellow, and white flesh (Figure 2C). Both pure and asymmetrical canary yellow individuals were classified into the yellow flesh group. In the F_2 population of 137 individuals, 106 plants showed yellow flesh and 31 showed white flesh, confirming a 3:1 Mendelian ratio ($\chi^2 = 0.199$, $p = 0.656$) (Table 1). In the F_3 population of 51 families, 40 families showed canary yellow or segregated flesh color and 11 families showed white flesh color, which also fitted the expected Mendelian ratio of 3:1 ($\chi^2 = 0.154$, $p = 0.699$). All 39 individuals in the BC_1P_1 population had canary yellow flesh. In the BC_1P_2 population of 41 individuals, 16 showed canary yellow flesh and 25 showed white flesh, fitting the Mendelian ratio of 1:1 ($\chi^2 = 0.988$, $p = 0.32$). Thus, the flesh color variation in this biparental population is controlled by a single nuclear locus.

To confirm whether the white flesh results from the malfunction of the canary yellow flesh locus, the canary yellow and white flesh parents were crossed with the orange flesh parent (CG149). Both F_1 hybrids (CG149 × HSH-F and CG149 × Sanbai) showed a canary yellow flesh color (Figure 2B). Thus, the white flesh line Sanbai complemented the malfunction in xanthophyll biosynthesis in the orange flesh line CG149. Hence, the genes involved in xanthophyll biosynthesis in the white flesh parent might function normally (Figure 2D). These results indicated that the white flesh trait in watermelon is controlled by a single recessive locus, i.e., *Clwf2*, and is epistatic to the canary yellow flesh.

Clwf2 is located on chromosome 6

For preliminary mapping of the white flesh locus, the SNP panel containing 390 SNP markers (Supplementary Table S1) was used to identify the genotype of the 137 F_2 individuals and the 51 F_3 families. Based on the phenotypic data of the F_2 population and the F_3 families, a locus associated with variations in flesh color was mapped to the

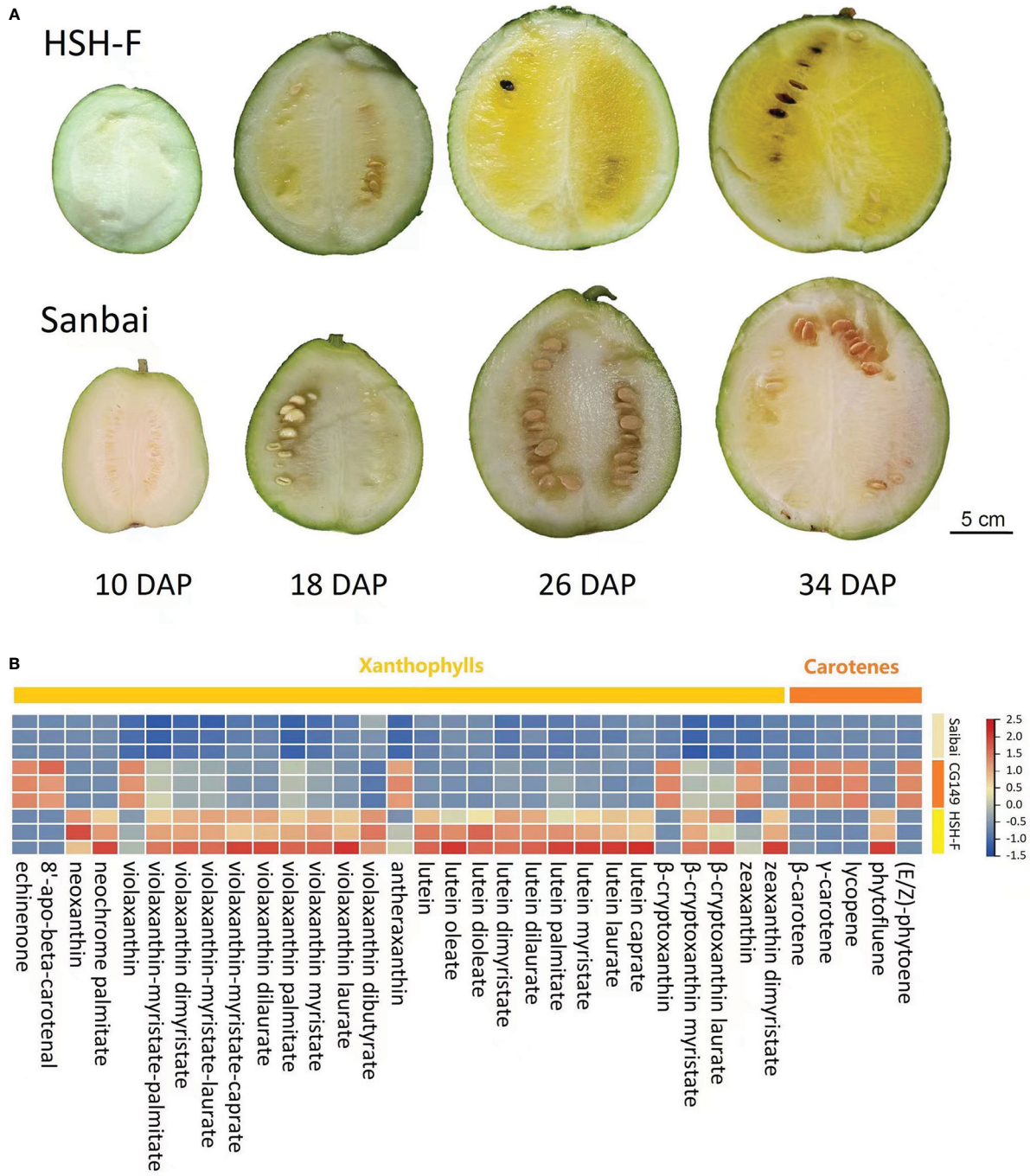


FIGURE 1
Characterization of the flesh color formation and carotenoid content of parental lines. **(A)** Fruits at 10, 18, 26, and 34 days after pollination (DAP) of HSH-F and Sanbai. **(B)** Heatmap of the carotenoid content of fruits from the parental lines at 34 DAP.

watermelon chromosome 6 in an 8.58-Mb region (Figure 3A) between the markers M401 (15.92 Mb) and M412 (24.50 Mb). The highest LOD scores for the 2019 F_2 and 2019 F_3 populations were 21.43 and 13.22, which explained 65.543% and 63.55% of the phenotypic variation, respectively (Supplementary Table S3).

To fine-map the *Clwf2* locus, the genomes of the parental lines HSH-F and Sanbai were resequenced using the Illumina HiSeq platform for marker development. A total of 96,524 variants were identified between the two parents, including 86,230 SNPs and 10,294 InDels (Supplementary Table S4). Firstly, six InDel markers were designed and used to identify the genotype of the 137 F_2 individuals used for preliminary mapping. This narrowed the *Clwf2* locus in the region of 1.02 Mb between markers ID_23467996 (23.47 Mb) and ID_24488295 (23.49 Mb) (Figure 3B). The markers ID_23467996 and ID_24488295 were then used to screen the recombinants in an F_2 population of 1,054 individuals at the seedling stage. Then, the recombinants were transplanted in the greenhouse for phenotyping and were further genotyped using seven SNP markers (Figure 3B).

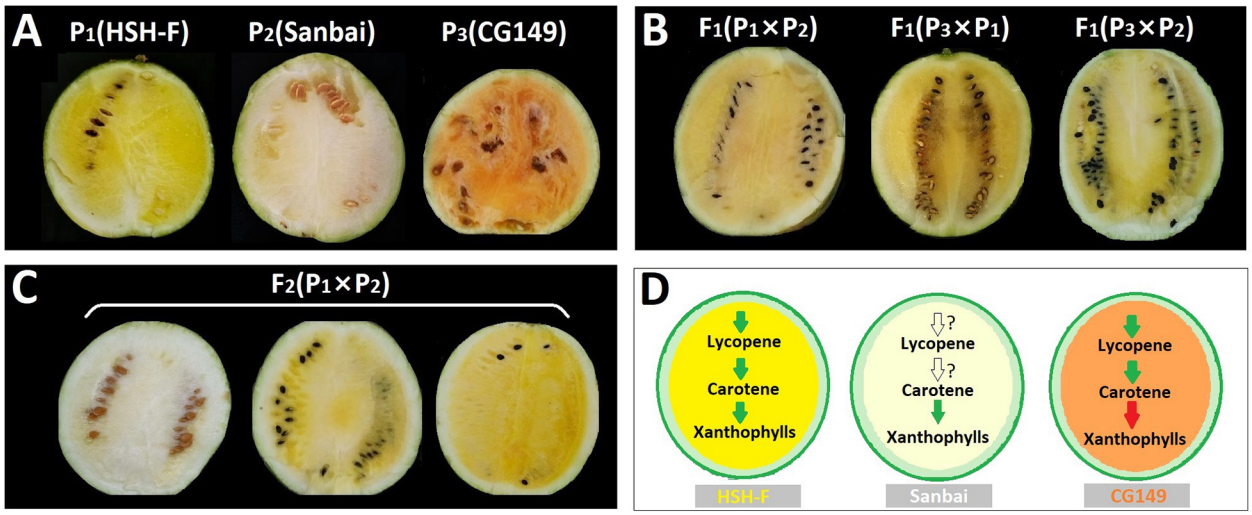


FIGURE 2
(A–C) Phenotype of the flesh colors of the parental lines **(A)**, F_1 hybrids **(B)**, and the F_2 segregation population **(C)** at 34 days after pollination (DAP). **(D)** Schematic of the key steps in the carotenoid biosynthesis pathway in the three parents. *Green arrows* indicate normal function, *red arrows* indicate function deficit, and *white arrows* indicate unknown function.

Finally, the *Clwf2* locus was delimited between the markers S_21474486 and S_24306830, with a physical distance of 132.3 kb (Chr06:24,174,486–24,306,830) (Figure 3B).

Several genes in the candidate region are downregulated in white flesh

According to the watermelon reference genome 97103 (v2), the 132.3-kb genome region contains 13 putative genes (Figure 3C; Supplementary Table S4), 11 of which have been annotated. However, no gene in this region was annotated in the CBP. The 11 annotated genes include two leucine-rich repeat receptor-like kinases (LRR-RLKs) (*Cla97C06G121800* and *Cla97C06G121810*), four chloroplastic proteins (*Cla97C06G121830*, *Cla97C06G121870*, *Cla97C06G121880*, and *Cla97C06G121890*), one cyclin-dependent kinase (CDK)-activating kinase assembly factor (*Cla97C06G121840*), one ATP-dependent DNA helicase (*Cla97C06G121850*), one APETALA2-like ethylene-responsive factor (AP2/ERF) (*Cla97C06G121910*), one heptahelical transmembrane protein (*Cla97C06G121900*), and one histone-lysine *N*-methyltransferase (*Cla97C06G121920*). For the four chloroplastic protein genes, *Cla97C06G121830* encodes a pentatricopeptide repeat-containing protein (PPR), *Cla97C06G121870* encodes a 50S ribosomal protein, *Cla97C06G121880* encodes a DNA-directed RNA polymerase subunit beta, and *Cla97C06G121890* encodes a tetratricopeptide repeat (TPR)-like superfamily protein.

No variant was identified in the candidate region in the comparison of the assembled genome resequencing data of the two parents. The expression levels of the 13 genes in the flesh samples from the two parents at 10, 18, 26, and 34 DAP were then analyzed using qRT-PCR. Among the 13 genes, four (*Cla97C06G121800*, *Cla97C06G121810*, *Cla97C06G121820*, and *Cla97C06G121870*) were only expressed in white flesh at 34 DAP (Supplementary Figure S2). Most of the other nine genes showed low or no expression at 10 DAP in both parents, but their expression levels progressively increased during fruit ripening (Figure 4). The expression levels of *Cla97C06G121860*, *Cla97C06G121880*, *Cla97C06G121890*, and *Cla97C06G121900* were significantly higher in the canary yellow than in the white flesh at 18

TABLE 1 Segregation ratio of the flesh color in the F_2, F_3, and BC_1 populations derived from crossing the white flesh with the canary yellow flesh inbred line (HSH-F × Sanbai).

Generation	Total	Yellow	White	Expected (observed) segregation ratio	χ^2 value	*p*-value
P_1 (HSH-F)	10	10	0			
P_2 (Sanbai)	8	0	8			
F_1 ($P_1 \times P_2$)	6	6	0			
F_2	137	106	31	3:1 (0.77:0.23)	0.199	0.656
$F_{2:3}$	51	40[a]	11[b]	3:1 (0.78:0.22)	0.154	0.699
BC_1P_1	39	39	0	–		–
BC_1P_2	41	16	25	1:1 (0.39:0.61)	0.988	0.32

[a]Fifteen individuals for each family, including all yellow flesh color or segregated flesh families
[b]All white flesh families.
χ^2, Chi-square test.

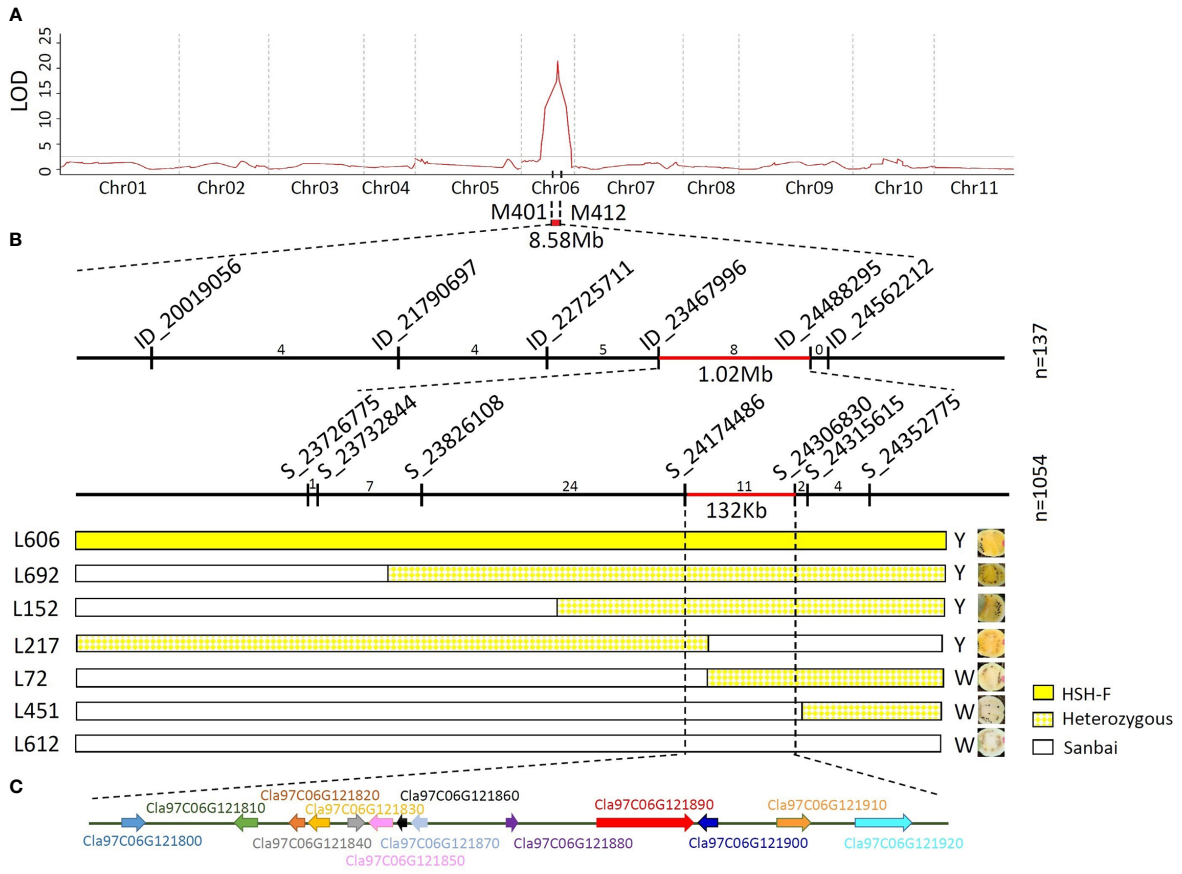


FIGURE 3
Genetic mapping of the *Clwf2* locus. **(A)** Preliminary mapping of the *Clwf2* locus. **(B)** Fine mapping of the *Clwf2* locus. **(C)** Genes in the candidate region for *Clwf2*. *Numbers above the chromosome* represent recombinants between adjacent markers; *n* is the number of plants used for mapping.

and 26 DAP. Notably, *Cla97C06G121890* and *Cla97C06G121900* showed almost no expression in immature white flesh fruits (10, 18, and 26 DAP). *Cla97C06G121890* is closely related to the *Arabidopsis* (*Arabidopsis thaliana*) REDUCED CHLOROPLAST COVERAGE (REC) proteins, which contribute to establishing the size of the chloroplast compartment and are essential for pigmentation (Larkin et al., 2016). Thus, the white flesh trait in watermelon might be related to the downregulation of the chloroplastic protein.

Plastid division genes are downregulated in white flesh

Chromoplasts are plastids that accumulate the carotenoid pigments in plants (Egea et al., 2010; Li and Yuan, 2013; Sun et al., 2018). Studies have demonstrated that chloroplasts and chromoplasts are interconvertible (Egea et al., 2010; Li and Yuan, 2013). *RCP2*, a homolog of *Cla97C06G121890* in *Mimulus*, was required for chromoplast development (Stanley et al., 2020). To investigate whether the repressed expression of the chloroplast protein gene affects the development of chromoplasts in watermelon, we examined the ultrastructure of the chromoplasts in flesh cells by TEM at 34 DAP. Notably, rounded and regularly shaped chromoplasts were observed in both canary yellow and white flesh cells (Figures 5A–D).

The chromoplast number and size have been reported to affect the accumulation of carotenoids (Sun et al., 2020). We subsequently assessed chromoplast proliferation in white and canary yellow flesh. Because the flesh cell of watermelon is too large to calculate the accurate number of chromoplasts, we then evaluated the relative expression levels of the plastid division genes in fruits of the white and canary yellow flesh parents, including *ClFtsZ1*, *ClFtsZ2*, *ClARC1*, *ClARC6*, *ClCDP*, *ClPDV1*, and *ClPDV2* (Osteryoung and Pyke, 2014; Chen et al., 2019). Most of the chloroplast proliferation genes were significantly downregulated during white flesh fruit development, particularly the *FtsZ* genes (Figure 5E). Thus, chromoplast proliferation may be repressed in white flesh watermelon.

The carotenoid biosynthesis pathway is downregulated in white flesh

The C20 GGPP molecule is the direct precursor for carotenoid biosynthesis. We first examined the relative expression level of *GGPPS* (*Cla97C02G039100*) in the fruits of the canary yellow and white flesh parents. The expression level of *GGPPS* in canary yellow flesh was significantly higher than that in white flesh during the whole fruit development stage (Figure 6). *PSY* is the primary bottleneck gene in carotenoid biosynthesis (Nisar et al., 2015; Yuan et al., 2015). Three *PSY*

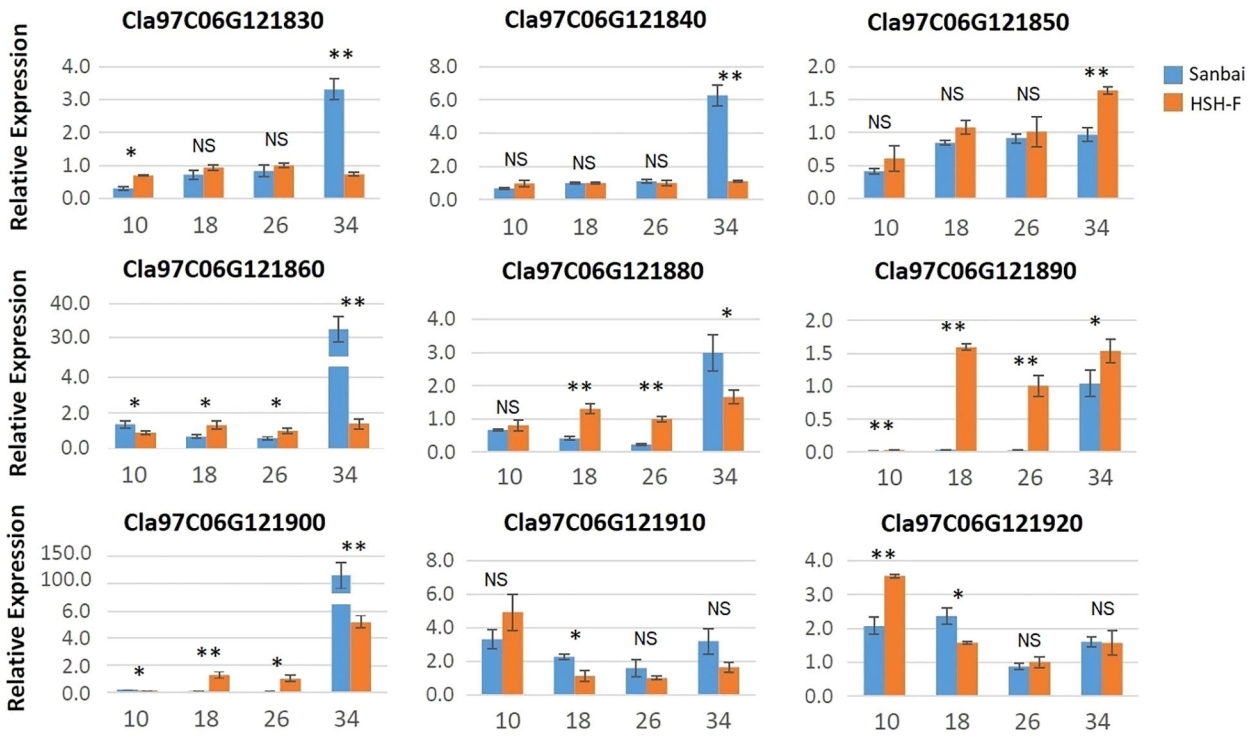


FIGURE 4
Relative expression of the candidate genes in fruits at different developmental stages of HSH-F and Sanbai. *Numbers on the x*-axis—10, 18, 26, and 34—represent fruits at 10, 18, 26, and 34 days after pollination (DAP), respectively. *Error bars* indicate the mean ± SD (n = 3). Significant difference between the two parents at each stage was assessed using Student's t-test. *p < 0.05; **p < 0.01. NS, no significance.

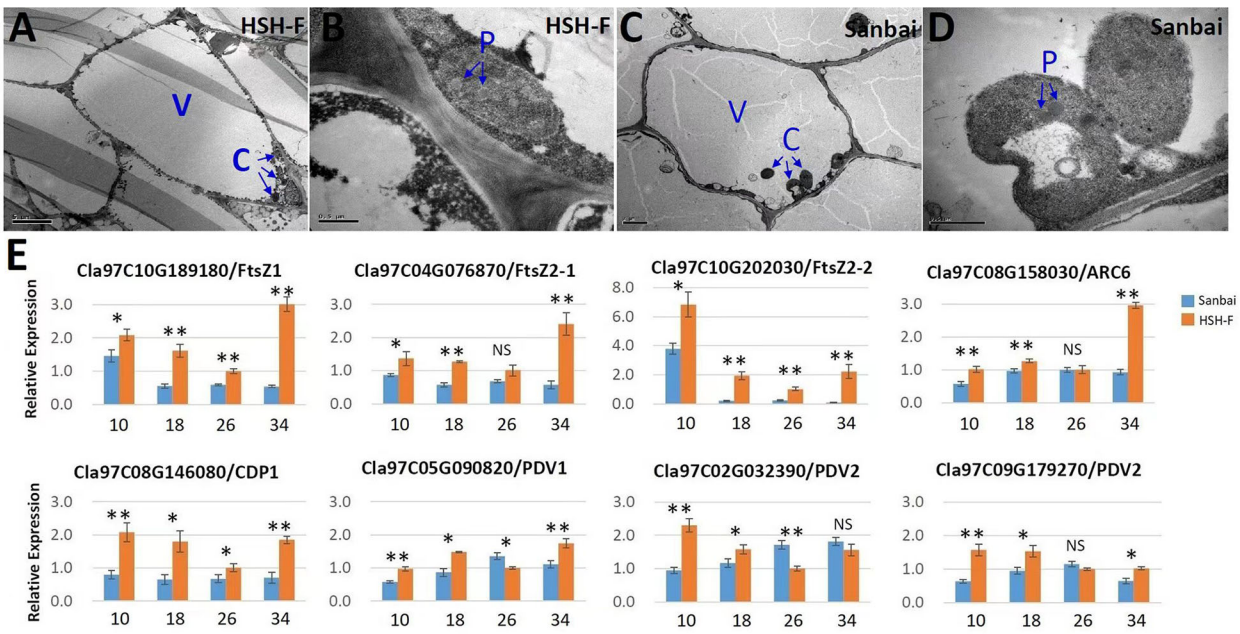


FIGURE 5
(A–E) Transmission electron microscopy of chromoplasts **(A–D)** and relative expression of the chromoplast proliferation-related genes **(E)**. **(A, B)** Micrographs of HSH-F fruit cells at 34 days after pollination (DAP): whole-cell view **(A)** and detailed view of chromoplasts **(B)**. **(C, D)** Micrographs of Sanbai fruit cells at 34 DAP: whole-cell view **(C)** and detailed view of chromoplasts **(D)**. *V*, vacuole; *C*, chromoplast; *P*, plastoglobule. *Error bars* indicate the mean ± SD (n = 3). Significant difference between the two parents at each stage was assessed using Student's t-test. *p < 0.05; **p < 0.01. NS, no significance.

10.3389/fpls.2023.1090009

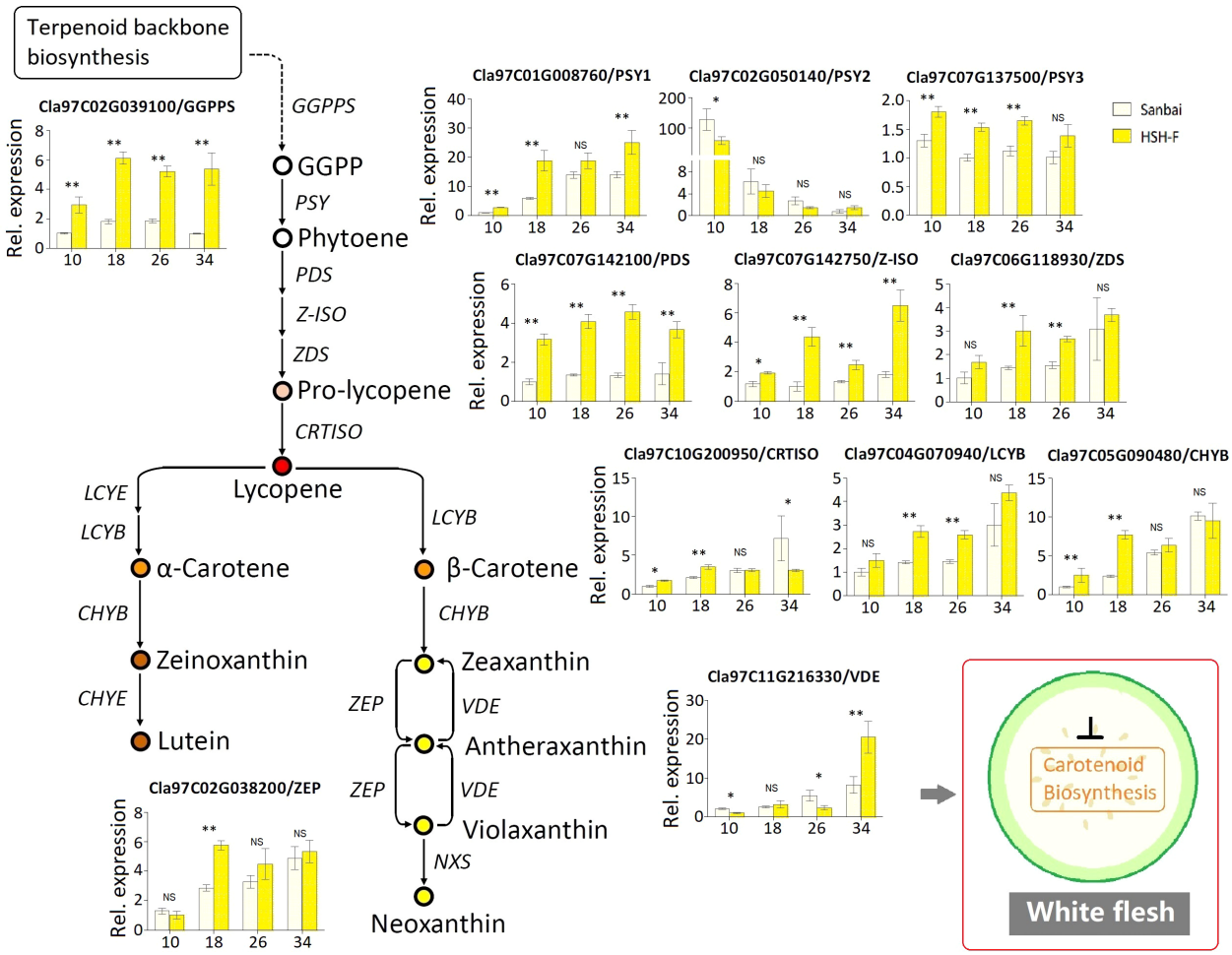


FIGURE 6
Relative expression of the carotenoid biosynthesis pathway genes in fruits of the canary yellow and white flesh parents. *Numbers on the x*-axis—10, 18, 26, and 34—represent fruits at 10, 18, 26, and 34 days after pollination (DAP), respectively. *PSY*, phytoene synthase; *GGPPS*, geranylgeranyl pyrophosphate synthase; *PDS*, phytoene desaturase; *Z-ISO*, 15-*cis*-ζ-carotene isomerase; *ZDS*, ζ-carotene desaturase; *CRTISO*, lycopene isomerase; *LCYB*, lycopene *b*-cyclase; *CHYB*, carotenoid β-hydroxylase; *ZEP*, zeaxanthin epoxidase; *VDE*, violaxanthin de-epoxidase. *Error bars* indicate the mean ± SD (n = 3). Significant difference between the two parents at each stage was assessed using Student's *t*-test. *p < 0.05; **p < 0.01. NS, no significance.

genes have been annotated in the watermelon genome: *PSY1* (*Cla97C01G008760*), *PSY2* (*Cla97C02G050140*), and *PSY3* (*Cla97C07G137500*). *PSY1* had a much higher expression level in watermelon fruits compared to *PSY2* and *PSY3* (Figure 6). The expression level of *PSY1* gradually increased with fruit ripening, and *PSY1* showed a significantly higher expression level in canary yellow flesh than in white flesh at all developmental stages. Similarly, the *PDS* (*Cla97C07G142100*) and *Z-ISO* (*Cla97C07G142750*) genes showed significantly higher expression levels in canary yellow flesh than in white flesh at all developmental stages. We also examined the expression levels of *ZDS* (*Cla97C06G118930*), *CRTISO* (*Cla97C10G200950*), *LCYB* (*Cla97C04G070940* and *Cla97C10G193770*), *CHYB* (*Cla97C05G090480*), and *ZEP* (*Cla97C02G038200*). These genes exhibited increased expression levels with fruit ripening in both parental lines; however, all genes showed significantly higher expression levels in the canary yellow flesh than in the white flesh at 18 or 26 DAP, the vital stages for pigmentation (Figure 6). Thus, the entire CBP is repressed in white flesh.

Discussion

Carotenoids are important bioactive metabolites that contribute to the beautiful colors of yellow, orange, and red in vegetables, fruits, and flowers and improve the nutritional quality in the human diet. The CBP has been well established, and the core genes and enzymes that catalyze carotenoid biosynthesis and degradation have been cloned and studied in many plants (Nisar et al., 2015; Yuan et al., 2015; Sun et al., 2018). However, regulation of the accumulation of carotenoids, which leads to the varied carotenoid content and composition among or within plant species, is not yet clear. Studies on major horticultural crops, such as tomato, citrus, and melon, have revealed diverse regulation mechanisms of carotenoid accumulation (Kato, 2012; Liu et al., 2015; Feder et al., 2019). Transcriptional regulation of the CBP genes is a major mechanism underlying carotenoid accumulation. Enhanced transcription of the bottleneck gene *PSY* along with *PDS* and *CRSTIO* increased the total carotenoid contents in tomato, watermelon, and red pepper (Grassi et al., 2013; Lv et al., 2015; Berry et al., 2019; Ilahy et al., 2019). A high transcript

abundance of *LCYB*/*LCYE* was associated with the high accumulation of carotene in fruits (Lu et al., 2006; Ampomah-Dwamena et al., 2009; Liu et al., 2015). In watermelon, the production of violaxanthin and lutein in yellow flesh was positively correlated with the expression levels of *CHYB* and *ZEP* (Lv et al., 2015; Yuan et al., 2021). In addition, the modulation of carotenogenic enzymes and the regulation of chromoplast biogenesis, development, and differentiation have been demonstrated to strongly affect carotenoid accumulation (Yuan et al., 2015; Sun et al., 2018; Feder et al., 2019). Despite great progress having been made in the research of carotenoid metabolism in plants, little is known about carotenoid regulation in watermelon. In this study, genetic mapping of the flesh color regulation gene was performed using a cross from yellow flesh and white flesh watermelon, and the probable mechanism underlying the white flesh color was elucidated.

Several flesh color loci have been identified in watermelon; however, the inheritance of flesh color, especially white flesh, is not yet clear. *Wf* is the first documented locus for the white flesh trait in watermelon. It is dominant epistatic to yellow flesh (*B*) and red flesh (*b*), as a segregation ratio of 12 white:3 yellow:1 red was obtained in the F_2 generation of a cross between the white and red flesh inbred lines (Shimotsuma, 1963). Another study using the F_2 population from a cross between the white (COS) and canary yellow (NC-517) flesh inbred lines suggested that the *C* gene for the canary yellow flesh color is dominant over the white flesh (Gusmini and Wehner, 2006). Recently, genetic inheritance analysis using two segregation populations from crossing Bingtangcui (white flesh) with Xihua (canary yellow flesh) and Sashengnaiyougua (white flesh) with Xinjinlanxuan (canary yellow flesh) has suggested that canary yellow is recessive to white in watermelon flesh (Diao et al., 2021). In this study, the flesh color inheritance analysis using the F_2 and F_3 segregation populations produced from the cross between the canary yellow (HSH-F) and white (Sanbai) flesh lines obtained a 3:1 segregation ratio of canary yellow/white flesh (Table 1). Consistent with previous studies, the canary yellow flesh contained abundant xanthophyll, whereas white flesh watermelon accumulated trace carotenoids (Figure 1B and Supplementary Table S2) (Bang et al., 2010; Zhao et al., 2013; Lv et al., 2015; Diao et al., 2021; Yuan et al., 2021). To examine whether the white flesh in Sanbai resulted from the malfunction of the canary yellow flesh locus, we crossed an orange flesh line (CG149), which accumulated massive carotene but lacked xanthophyll (Figure 1B and Supplementary Table S2), with the canary yellow (HSH-F) and white flesh (Sanbai) lines. Interestingly, both F_1 hybrids (CG149 × HSH-F and CG149 × Sanbai) had canary yellow flesh, similarly to the F_1 hybrid that resulted from crossing HSH-F with Sanbai (Figure 2B). This result indicated that both Sanbai and HSH-F complement the xanthophyll biosynthesis deficient in the orange flesh line CG149. Thus, the above findings suggest that the *Clwf2* locus for the white flesh color and the *C* locus for the canary yellow flesh are two genetic loci. Moreover, *Clwf2* is different from the *Wf* locus because the white flesh in this study is recessive epistatic to the canary yellow flesh.

The well-assembled genome sequence facilitated the genetic mapping and cloning of the function genes in watermelon. Three flesh color-related loci have been reported on chromosome 6 (Guo et al., 2019; Li et al., 2020; Diao et al., 2021). Li et al. (2020) narrowed a scarlet red flesh color locus (i.e., Y^{scr}) in a 40-kb region on chromosome 6 that contains five putative genes (*Cla018767*–*Cla018771*). Diao et al. (2021) located a canary yellow flesh locus in a 600-kb region (24.00–24.61 Mb) on chromosome 6. Based on functional annotation and qRT-PCR analysis, the genes *Cla97C06G121680*, *Cla97C06G121700*, *Cla97C06G121890*, *Cla97C06G122090*, and *Cla97C06G121910* were predicted as the candidate genes for the canary yellow flesh in watermelon. In this study, genetic mapping using the canary yellow (HSH-F) and white (Sanbai) flesh lines narrowed the white flesh locus *Clwf2* to a 132.3-kb region on chromosome 6 (Chr06:24,174,486–24,306,830) (Figure 3). According to the watermelon reference genome 97103 (v2), the candidate region for the *Clwf2* locus overlapped with the candidate regions for Y^{scr} and the canary yellow flesh locus. These findings emphasize the important role of this chromosome region in regulating the flesh color in watermelon. Notably, no CBP gene was identified in these candidate regions. The study by Ren et al. (2018) showed that overexpression of the tonoplast sugar transporter gene *ClTST2* (*Cla97C02G036390*) can lead to flesh color development in watermelon. Moreover, the watermelon fruit chromoplast-localized phosphate transporter ClPHT4;2 (*Cla97C10G205070*) was necessary for carotenoid accumulation in fruit flesh (Zhang et al., 2017). These studies concluded that non-CBP genes also play important roles in flesh color regulation in watermelon.

Many transcriptional regulators have been demonstrated to regulate the accumulation of carotenoids in vegetables, flowers, and fruits (Liu et al., 2015; Yuan et al., 2015; Stanley et al., 2020). For example, the MADS-box transcription factor (TF) RIPENING INHIBITOR (*RIN*), STAY-GREEN 1 (*SGR1*), AP2/ethylene response factor-type TF (*RAP2.2*), and phytochrome interacting factor 1 (*PIF1*) regulate the accumulation of carotenoids in tomato fruits by interacting with the *SlPSY1* promoter (Liu et al., 2015). In watermelon, four PPR family genes, which are important RNA-binding proteins in plants, were correlated with flesh color (Subburaj et al., 2020). In the present study, several transcriptional regulators were identified in the candidate region for *Clwf2*, such as the PPR protein (*Cla97C06G121830*), the TPR protein (*Cla97C06G121890*), and AP2/ERF (*Cla97C06G121910*) (Supplementary Table S5). Among the TFs, the TPR protein gene (*Cla97C06G121890*) showed almost no expression in immature fruits (10, 18, and 26 DAP) of white flesh watermelon (Figure 4). This result is consistent with the study by Diao et al. (2021), in which *Cla97C06G121890* had a much higher expression level in canary flesh than in white flesh watermelon. TPRs are known to mediate protein–protein interactions in diverse biological processes (Zeytuni and Zarivach, 2012). In *Arabidopsis*, the loss of function of the TPR protein gene (*REC*) decreased the chlorophyll content and caused smaller chloroplasts in leaves (Larkin et al., 2016). More recently, the TPR protein gene (*RCP2*) was found to be necessary and sufficient for chromoplast development and carotenoid accumulation in *Mimulus* floral tissues, and the loss-of-function mutation of *RCP2* caused the malformation of chromoplasts and the drastic downregulation of the entire CBP (Stanley et al., 2020). Similar to the study on *Mimulus* flowers, we found that almost all CBP genes were downregulated in white flesh watermelon fruits (Figure 6). Thus, the white flesh trait in watermelon is probably related to the repression of the TPR protein gene *Cla97C06G121890*. However, not only *Cla97C06G121890* but

also three other genes (*Cla97C06G121860*, *Cla97C06G121880*, and *Cla97C06G121900*) adjacent to *Cla97C06G121890* were downregulated in white flesh fruits compared to canary yellow flesh fruits (Figure 4). Thus, further studies investigating the mechanisms underlying transcriptional repression of this genomic region will help in understanding the formation of white flesh in watermelon.

Chromoplasts are the organelles for carotenoid biosynthesis and storage in plant cells, and their number and size are closely related to the accumulation of carotenoids (Li and Yuan, 2013; Sun et al., 2018). In addition to the TPR proteins, the *Or* gene encodes a plastid-associated protein that has been found to increase the accumulation of β-carotene in cauliflower, melon fruits, and tomato flower by triggering chromoplast differentiation (Paolillo et al., 2004; Lu et al., 2006; Chayut et al., 2017; Yazdani et al., 2019). In *Arabidopsis*, *Or* controlled the chromoplast number by specifically interacting with the chloroplast division regulator ACCUMULATION AND REPLICATION OF CHLOROPLASTS 3 (*ARC3*), and overexpression of the plastid division factor *PLASTID DIVISION 1* greatly enhanced the carotenoid accumulation in calli (Sun et al., 2020). Tomato mutants with enhanced plastid number and size (i.e., *hp1*, *hp2*, and *hp3*) had higher levels of carotenoids in fruits (Liu et al., 2004; Kolotilin et al., 2007; Galpaz et al., 2008). In the present study, rounded and regularly shaped chromoplasts were observed in both canary yellow and white flesh fruit cells (Figures 5A–D), indicating that the lack of carotenoid in white flesh watermelon did not result from the malformation of chromoplasts. However, the qRT-PCR analysis showed that several key genes associated with plastid division, such as the *FtsZ* genes (*Cla97C10G189180*, *Cla97C04G076870*, and *Cla97C10G202030*), were significantly downregulated in white flesh watermelon fruits (Figure 5E). Thus, whether chromoplast differentiation is inhibited in white flesh watermelon and subsequently blocks carotenoid accumulation is worth further study.

In conclusion, we characterized a *Clwf2* locus that determines the white flesh trait in watermelon. From the results of the carotenoid content determination, genetic mapping, expression level evaluation, and TEM analyses, we conclude that *Clwf2* regulates the white flesh trait probably by repressing the expression of the TPR protein gene (*Cla97C06G121890*), which further inhibits the proliferation of chromoplasts and causes downregulation of the CBP genes, blocking the accumulation of carotenoids.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession numbers SAMN22357443 (HSH-F, https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN22357443) and SAMN22357441 (Sanbai, https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN22357441).

Author contributions

ZD and LY designed the study. LY, WZ, ZC, and NW contributed to the experiments. LY and YZ analyzed the experimental data. ZD provided the seeds for the experiment. LY wrote the manuscript. YZ and YW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (32002065), the Key R&D project of Hubei province (2020BBA037), the Natural Science Foundation of Hubei province (2021CFB390), The Earmarked Fund for China Agriculture Research System (CARS-25), the Innovation Team Project of Hubei Agricultural Science and Technology Innovation Center (2021-620-000-001-43201), and the Young Foundation of Hubei Academy of Agricultural Sciences (2020NKYJJ05).

Acknowledgments

We wish to thank all the reviewers and editors for their careful reading and helpful comments on this manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1090009/full#supplementary-material

References

Ampomah-Dwamena, C., McGhie, T., Wibisono, R., Montefiori, M., Hellens, R., and Allan, A. (2009). The kiwifruit lycopene beta-cyclase plays a significant role in carotenoid accumulation in fruit. *J. Exp. Bot.* 60 (13), 3765–3779. doi: 10.1093/jxb/erp218

Ampomah-Dwamena, C., Thrimawithana, A., Dejnoprat, S., Lewis, D., Espley, R., and Allan, A. (2019). A kiwifruit (Actinidia deliciosa) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* 221, 309–325. doi: 10.1111/nph.15362

Bang, H., Davis., A., Kim, S., Leskovar, D., and King, S. (2010). Flesh color inheritance and gene interactions among canary yellow, pale yellow, and red watermelon. *J. Amer Soc. Hort Sci.* 135, 362–368. doi: 10.21273/JASHS.135.4.362

Bang, H., Kim, S., Leskovar, D., and King, S. (2007). Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β-cyclase (LCYB) gene. *Mol. Breed.* 20, 63–72. doi: 10.1007/s11032-006-9076-4

Berry, H., Rickett, D., Baxter, C., Enfissi, E., and Fraser, P. (2019). Carotenoid biosynthesis and sequestration in red chilli pepper fruit and its impact on colour intensity traits. *J. Exp. Bot.* 70 (10), 2637–2650. doi: 10.1093/jxb/erz086

Branham, S., Vexler, L., Meir, A., Tzuri, G., Frieman, Z., Levi, A., et al. (2017). Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. *Mol. Breed.* 37, 146. doi: 10.1007/s11032-017-0747-0

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

Chayut, N., Yuan, H., Ohali, S., Meir, A., Sa'ar, U., Tzuri, G., et al. (2017). Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. *Plant Physiol.* 173, 376–389. doi: 10.1104/pp.16.01256

Chen, C., Cao, L., Yang, Y., Porter, K. J., and Osteryoung, K. W. (2019). ARC3 activation by PARC6 promotes FtsZ-ring remodeling at the chloroplast division site. *Plant Cell* 31, 862–885. doi: 10.1105/tpc.18.00948

Diao, W., Yuan., P., Gong., C., Zhao., S., Zhu., H., and Lu., X. (2021). Genetic analysis and gene mapping of canary yellow in watermelon flesh. *Scientia Agricultura Sin.* 54 (18), 3945–3958. doi: 10.3864/j.issn.0578-1752.2021.18.013

Duan, X., Cai., J., Zhao, Y., Gao, G., Li, M., and Qi, H. (2022). Transcriptome and metabolomics analysis revealed that CmWRKY49 regulating CmPSY1 promotes b-carotene accumulation in orange fleshed oriental melon. *Hortic. Plant J.* 8, 650e666. doi: 10.1016/j.hpj.2022.07.005

Egea, I., Barsan, C., Bian, W., Purgatto, E., Latche, A., Chervin, C., et al. (2010). Chromoplast differentiation: current status and perspectives. *Plant Cell Physiol.* 51, 1601–1611. doi: 10.1093/pcp/pcq136

Feder, A., Chayut, N., Gur, A., Freiman, Z., Tzuri, G., Meir, A., et al. (2019). The role of carotenogenic metabolic flux in carotenoid accumulation and chromoplast differentiation: Lessons from the melon fruit. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01250

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

Fu, C., Han, Y., Kuang, J., Chen, J., and Lu, W. (2017). Papaya CpEIN3a and CpNAC2 Co-operatively regulate carotenoid biosynthesis-related genes CpPDS2/4, CpLCY-e and CpCHY-b during fruit ripening. *Plant Cell Physiol.* 58, 2155–2165. doi: 10.1093/pcp/pcx149

Galpaz, N., Wang, Q., Menda, N., Zamir, D., and Hirschberg, J. (2008). Abscisic acid deficiency in the tomato mutant high-pigment 3 leading to increased plastid number and higher fruit lycopene content. *Plant J.* 53 (5), 717–730. doi: 10.1111/j.1365-313X.2007.03362.x

Grassi, S., Piro, G., Lee, J., Zheng, Y., Fei, Z., Dalessandro, G., et al. (2013). Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. *BMC Genomics* 14, 781. doi: 10.1186/1471-2164-14-781

Guo, S., Zhang, J., Sun., H., Salse, J., Lucas, W., Zhang, H., et al. (2012). The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat. Genet.* 45 (1), 51–58. doi: 10.1038/ng.2470

Guo, S., Zhao., S., Sun, H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4

Gusmini, G., and Wehner, T. (2006). Qualitative inheritance of rind pattern and flesh color in watermelon. *J. Hered* 97, 177–185. doi: 10.1093/jhered/esj023

Henderson, W. (1989). Inheritance of orange flesh color in watermelon. *Cucurbit Genet. Coop Rep.* 12, 59–63.

Henderson, W., Scott, G., and Wehner, T. (1998). Interaction of flesh color genes in watermelon. *J. Hered* 89, 50–53. doi: 10.1093/jhered/89.1.50

Ilahy, R., Tlili, I., Siddiqui, M., Hdider, C., and Lenucci, M. (2019). Inside and beyond color: Comparative overview of functional quality of tomato and watermelon fruits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00769

Kato, M. (2012). Mechanism of carotenoid accumulation in citrus fruit. *J. Japanese Soc. Hortic. Sci.* 81, 219–233. doi: 10.2503/JJSHS1.81.219

Kolotilin, I., Koltai, H., Tadmor, Y., Bar-Or, C., Reuveni, M., Meir, A., et al. (2007). Transcriptional profiling of high pigment-2dg tomato mutant links early fruit plastid biogenesis with its overproduction of phytonutrients. *Plant Physiol.* 145 (2), 389–401. doi: 10.1104/pp.107.102962

Larkin, R., Stefano, G., Ruckle, M., Stavoe, A., Sinkler, C., Brandizzi, F., et al. (2016). REDUCED CHLOROPLAST COVERAGE genes from arabidopsis thaliana help to establish the size of the chloroplast compartment. *Proc. Natl. Acad. Sci.* 113 (8), E1116–E1125. doi: 10.1073/pnas.1515741113

Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361–374. doi: 10.1534/genetics.106.066811

Li, L., and Yuan, H. (2013). Chromoplast biogenesis and carotenoid accumulation. *Arch. Biochem. Biophys.* 539, 102–109. doi: 10.1016/j.abb.2013.07.002

Li, N., Shang, J., Wang, J., Zhou, D., Li, N., and Ma, S. (2020). Discovery of the genomic region and candidate genes of the scarlet red flesh color (Y^{scr}) locus in watermelon (*Citrullus lanatus* l.). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00116

Liu, S., Gao, Z., Wang, X., Luan, F., Dai, Z., Yang, Z., et al. (2022). Nucleotide variation in the phytoene synthase (ClPsy1) gene contributes to golden flesh in watermelon (*Citrullus lanatus* l.). *Theor. Appl. Genet.* 135, 185–200. doi: 10.1007/s00122-021-03958-0

Liu, Y., Roof, S., Ye, Z., Barry, C., Tuinen, A., Vrebalov, J., et al. (2004). Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. *Proc. Natl. Acad. Sci.* 101 (26), 9897–9902. doi: 10.1073/pnas.0400935101

Liu, L., Shao, Z., Zhang, M., and Wang, Q. (2015). Regulation of carotenoid metabolism in tomato. *Mol. Plant* 8, 28–39. doi: 10.1016/j.molp.2014.11.006

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

Lu, S., Van Eck, J., Zhou, X., Lopez, A., O'Halloran, D., Cosman, K., et al. (2006). The cauliflower or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. *Plant Cell* 18, 3594–3605. doi: 10.1105/tpc.106.046417

Lu, S., Zhang, Y., Zhu, K., Yang, W., Ye, J., Chai, L., et al. (2018). The citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. *Plant Physiol.* 176, 2657–2676. doi: 10.1104/pp.17.01830

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

Ma, N., Feng, H., Meng, X., Li, D., Yang, D., Wu, C., et al. (2014). Overexpression of tomato SlNAC1 transcription factor alters fruit pigmentation and softening. *BMC Plant Biol.* 14, 351. doi: 10.1186/s12870-014-0351-y

Meng, Y., Wang, Z., Wang, Y., Wang, C., Zhu, B., Liu, H., et al. (2019). The MYB activator WHITE PETAL1 associates with MtTT8 and MtWD40-1 to regulate carotenoid-derived flower pigmentation in medicago truncatula. *Plant Cell* 31, 2751–2767. doi: 10.1105/tpc.19.00480

Nisar, N., Li, L., Lu, S., Khin, N., and Pogson, B. (2015). Carotenoid metabolism in plants. *Mol. Plant* 8, 68–82. doi: 10.1016/j.molp.2014.12.007

Osteryoung, K., and Pyke, K. (2014). Division and dynamic morphology of plastids. *Annu. Rev. Plant Biol.* 65, 443–472. doi: 10.1146/annurev-arplant-050213-035748

Paolillo, J., Garvin, D., and Parthasarathy, M. (2004). The chromoplasts of Or mutants of cauliflower (Brassica oleracea L. var. botrytis). *Protoplasma* 224 (3-4), 245–53. doi: 10.1007/s00709-004-0059-1

Poole, C. (1944). Genetics of cultivated cucurbits. *J. Hered* 35, 122–128. doi: 10.1093/oxfordjournals.jhered.a105364

Porebski, S., Bailey, L., and Baum., B. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/bf02772108

Porter, D. (1937). Inheritance of certain fruit and seed characters in watermelons. *Hilgardia* 10, 489–509. doi: 10.3733/hilg.v10n12p489

Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* 176, 836–850. doi: 10.1104/pp.17.01290

Sagawa, J., Stanley, L., LaFountain, A., Frank, H., Liu, C., and Yuan, Y. (2016). An R2R3-MYB transcription factor regulates carotenoid pigmentation in mimulus lewisii flowers. *New Phytol.* 209, 1049–1057. doi: 10.1111/nph.13647

Shimotsuma, M. (1963). Cytogenetical studies in the genus citrullus. VI. inheritance of several characters in watermelons. *Jap J. Breed* 13, 235–240. doi: 10.1270/jsbbs1951.13.235

Stanley, L., Ding, B., Sun, W., Mou, F., Hill, C., Chen, S., et al. (2020). A tetratricopeptide repeat protein regulates carotenoid biosynthesis and chromoplast development in monkeyflowers (*Mimulus*). *Plant Cell* 32, 1536–1555. doi: 10.1105/tpc.19.00755

Subburaj, S., Tu, L., Lee, K., Park, G. S., Lee, H., Chun, J., et al. (2020). A genome-wide analysis of the pentatricopeptide repeat (PPR) gene family and PPR-derived markers for flesh color in watermelon (*Citrullus lanatus*). *Genes (Basel)* 11 (10), 1125. doi: 10.3390/genes11101125

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

Sun, T., Yuan, H., Chen, C., Kadirjan-Kalbach, D., Mazourek, M., Osteryoung, K., et al. (2020). OR(His), a natural variant of OR, specifically interacts with plastid division factor ARC3 to regulate chromoplast number and carotenoid accumulation. *Mol. Plant* 13 (6):864-878. doi: 10.1016/j.molp.2020.03.007

Wang, C., Luan, F., Liu, H., Davis, A., Zhang, Q., Dai, Z., et al. (2021). Mapping and predicting a candidate gene for flesh color in watermelon. *J. Integr. Agr* 20, 2100–2111. doi: 10.1016/S2095-3119(20)63487-6

Welsch, R., Maass, D., Voegel, T., Dellapenna, D., and Beyer, P. (2007). Transcription factor RAP2.2 and its interacting partner SINAT2: stable elements in the carotenogenesis of arabidopsis leaves. *Plant Physiol.* 145, 1073–1085. doi: 10.1104/pp.107.104828

Xiong, C., Luo, D., Lin, A., Zhang, C., Shan, L., He, P., et al. (2019). A tomato b-box protein SlBBX20 modulates carotenoid biosynthesis by directly activating PHYTOENE SYNTHASE 1, and is targeted for 26S proteasome-mediated degradation. *New Phytol.* 221, 279–294. doi: 10.1111/nph.15373

Yazdani, M., Sun, Z., Yuan, H., Zeng, S., Thannhauser, T. W., Vrebalov, J., et al. (2019). Ectopic expression of ORANGE promotes carotenoid accumulation and fruit development in tomato. *Plant Biotechnol. J.* 17 (1), 33–49. doi: 10.1111/pbi.12945

Yuan, P., Umer, M., He, M., Zhao, S., Lu, X., Zhu, H., et al. (2021). Transcriptome regulation of carotenoids in five flesh-colored watermelons (*Citrullus lanatus*). *BMC Plant Biol.* 21, 203. doi: 10.1186/s12870-021-02965-z

Yuan, H., Zhang, J., Nageswaran, D., and Li, L. (2015). Carotenoid metabolism and regulation in horticultural crops. *Hortic. Res.* 2, 15036. doi: 10.1038/hortres.2015.36

Zeytuni, N., and Zarivach, R. (2012). Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* 20, 397–405. doi: 10.1016/j.str.2012.01.006

Zhang, J., Guo, S., Ren, Y., Zhang, H., Gong, G., Zhou, M., et al. (2017). High-level expression of a novel chromoplast phosphate transporter ClPHT4;2 is required for flesh color development in watermelon. *New Phytol.* 213, 1208–1221. doi: 10.1111/nph.14257

Zhao, W., Lv, P., and Gu, H. (2013). Studies on carotenoids in watermelon flesh. *Agric. Sci.* 04, 13–20. doi: 10.4236/as.2013.47A003

Zhou, D., Shen, Y., Zhou, P., Fatima, M., Lin, J., Yue, J., et al. (2019). Papaya CpbHLH1/2 regulate carotenoid biosynthesis-related genes during papaya fruit ripening. *Hortic. Res.* 6, 80. doi: 10.1038/s41438-019-0162-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/fpls.2023.1137126

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Yunsong Lai,
Sichuan Agricultural University, China
Fuyou Fu,
Agriculture and Agri-Food Canada (AAFC), Canada
Yu Pan,
Southwest University, China

*CORRESPONDENCE
Yongdong Sun
ydsun2019@126.com

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

RECEIVED 04 January 2023
ACCEPTED 13 February 2023
PUBLISHED 23 February 2023

CITATION
Luo W, Zhao Z, Chen H, Ao W, Lu L, Liu J, Li X and Sun Y (2023) Genome-wide characterization and expression of *DELLA* genes in *Cucurbita moschata* reveal their potential roles under development and abiotic stress.
Front. Plant Sci. 14:1137126.
doi: 10.3389/fpls.2023.1137126

Genome-wide characterization and expression of *DELLA* genes in *Cucurbita moschata* reveal their potential roles under development and abiotic stress

Weirong Luo[1,2], Zhenxiang Zhao[1,2], Hongzhi Chen[3], Wenhong Ao[1,2], Lin Lu[1,2], Junjun Liu[1,2], Xinzheng Li[1,2] and Yongdong Sun[1,2]*

[1]School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang, China, [2]Henan Province Engineering Research Center of Horticultural Plant Resource Utilization and Germplasm Enhancement, Xinxiang, China, [3]College of Bioengineering, Xinxiang Institute of Engineering, Xinxiang, China

DELLA gene family plays a key role in regulating plant development and responding to stress. Currently, many *DELLA* family members have been identified in plants, however, information on *DELLA* genes in pumpkin (*Cucurbita moschata*) is scarce. In this study, physical and chemical properties, gene structure *cis*-regulatory elements and expression of *CmoDELLA* genes were examined in pumpkin. We found that seven *CmoDELLA* genes were identified in pumpkin, and they were unevenly classified into five chromosomes. CmoDELLA proteins were relatively unstable and their secondary structures were mainly made up α-helix and random coil. All seven CmoDELLA proteins contained typical DELLA domain and GRAS domain, however, motif numbers between CmoDELLA proteins were unevenly distributed, implying the complex evolution and functional diversification of CmoDELLA proteins. *Cis*-regulatory elements analysis revealed that *CmoDELLA* genes might play an essential role in regulating plant growth and development, and response to stress in pumpkin. Transcriptome data in the roots, stems, leaves and fruits demonstrated that *CmoDELLA2*, *CmoDELLA3* and *CmoDELLA7* were related to the stems development, *CmoDELLA1*, *CmoDELLA4*, *CmoDELLA5* and *CmoDELLA6* were associated with the fruits development. Furthermore, we found that *CmoDELLA1* and *CmoDELLA5* were up-regulated under NaCl stress. *CmoDELLA1*, *CmoDELLA2*, *CmoDELLA3*, *CmoDELLA5*, *CmoDELLA6* and *CmoDELLA7* were remarkably induced under waterlogging stress. While, all of the 7 *CmoDELLA* genes showed significantly induced expression under cold stress. The expression patterns under abiotic stress suggested that *CmoDELLA* genes might mediate the stress response of pumpkin to NaCl, waterlogging and cold, however, the functions of different *CmoDELLA* genes varied under different stress. Overall, our study provides valuable information for further research about the potential functions and regulatory networks of *CmoDELLA* genes in pumpkin.

KEYWORDS

***Cucurbita moschata*, *DELLA*, characterization, gene expression, abiotic stress**

1 Introduction

DELLA proteins play an essential role in regulating plant development and stress response, as key negative regulators of GA signaling (Zhou et al., 2020). DELLA protein sequences usually contain two conserved domains: DELLA domain and GRAS domain, which are distributed on the N-terminal region and C-terminal region, respectively. DELLA domain is known as the sensing domain of GA signaling for binding to GID1. And GRAS domain contributes mainly to repress GA responses by interacting with transcription factors (TFs), which is crucial for maintaining the functions of DELLA proteins (Xue et al., 2022). DELLA proteins do not contain the typical DNA-binding domain, however, they can interact with many TFs, such as PIFs, BZR1, EXP2, DREB1B, JAZs, and TCPs, to participate in almost all the processes of plant development and stress response (Thines et al., 2007; Navarro et al., 2008; Li et al., 2012; Li et al., 2016; Sechet et al., 2016; Liang et al., 2019). Identification of DELLA proteins firstly began in *Arabidopsis*, and AtRGL1, AtGAI, AtRGL2, AtRGA and AtRGl3, have been found as the members of the *Arabidopsis* DELLA family (Chen et al., 2013). DELLA proteins were subsequently identified from pear (Liu H. et al., 2016), strawberry (Li W.J. et al., 2018), cassava (Li X.L. et al., 2018), *Camellia sinensis* (Han et al., 2020), litchi (Wang et al., 2020), *Brassica napus* (Sarwar et al., 2021), Chinese cabbage (Guan et al., 2021) and other species.

DELLA genes play vital roles in regulating seed germination, hypocotyl elongation, plant height, flowering, fruit quality and stress response (Xue et al., 2022). Some studies have shown that *DELLA* genes can control seed germination by regulating the expression of multiple protein kinase genes (Cao et al., 2006). For instance, ABI3 and ABI5 can interact with DELLA proteins to activate SOMNUS and downstream target genes under high temperature, thereby inhibiting seed germination in *Arabidopsis* (Lim et al., 2013). DELLA proteins interacting with ARF6 and PIFs regulate cell elongation of the hypocotyl, which leading to short hypocotyl in *Arabidopsis* (Feng et al., 2008; Oh et al., 2014; Liu et al., 2018). Ethylene interacts with DELLA proteins to inhibit root growth and maintain apical hook-like structure of *Arabidopsis* (Achard et al., 2003). Furthermore, DELLA proteins interact with TCP to influence plant height through regulation of inflorescence apex growth (Davière et al., 2014), and interact with MONOCULM1 to affect tiller number and plant height of rice (Liao et al., 2019). DELLA proteins can also delay the floral transition by interacting with SPLs under long day conditions (Wang et al., 2009; Wu et al., 2009), and repress flowering by inhibiting *LFY* and *SOC1* genes expression under short day conditions (Achard et al., 2007). In addition, Fruit development can be regulated by DELLA proteins. Researchers have concluded that DELLA proteins are implicated in the fruit initiation by interacting with SlARF7/SlIAA9 (Hu et al., 2018), and silencing of *DELLA* gene results in parthenocarpic fruit in tomato (Martí et al., 2007). According to previous studies, DELLA proteins participate in abiotic stress response and improve the plant survival by regulating reactive oxygen species (ROS) levels during adverse environments (Achard et al., 2008a; Achard et al., 2008b). Study has shown that DELLA protein enhances the salt tolerance of wheat seedling by increasing superoxide dismutase (SOD) activity under salt stress (Wang et al., 2016).

Pumpkin (*Cucurbita moschata*) is an annual vegetable crop, which has edible, medicinal and ornamental values. Pumpkin is widely cultivated across the globe, and the top producer is China, which produced 7.7 million tonnes of pumpkin (Worldmapper, 2021). Pumpkin is also widely used as a grafting rootstock for other cucurbit vegetable crops, including cucumber, watermelon and melon, which promotes plant growth and strengthens biotic and abiotic stress tolerance (Liu S. S. et al., 2016; Li et al., 2017; Zhang et al., 2019). In recent years, during cultivation, extreme weather and unfavorable environment conditions such as inappropriate temperature, drought stress, waterlogging stress and salt stress, seriously limited the growth and development of cucurbit vegetable crops, leading to the decline of yield and quality. Therefore, it is of significant importance to elucidate the stress response mechanism of pumpkin and screen pumpkin resistant rootstock for strengthening the stress resistance of cucurbit vegetable crops. However, there is no report of *DELLA* gene family in pumpkin development and stress response.

This main purpose of this study is to identify and characterize *DELLA* gene family in pumpkin, and to uncover their potential functions under abiotic stress. In this current study, seven *CmoDELLA* genes were identified in pumpkin. Moreover, the chromosomal localization, protein properties, phylogenetic analysis, gene structure, promoter *cis*-regulatory elements, protein interaction networks and expression of *CmoDELLA* genes were investigated. Additionally, physiological changes of pumpkin seedlings under waterlogging stress and cold stress were detected. This is the first report on the genome-wide characterization and expression of *CmoDELLA* genes in pumpkin, which provides valuable clues on the biological functions of *CmoDELLA* genes in regulating plant development and stress response for further research.

2 Materials and methods

2.1 Identification of CmoDELLA family members in pumpkin

Pumpkin protein sequences were downloaded from the cucurbit genomics database (http://cucurbitgenomics.org/). Hidden Markov Model (HMM) profile of the DELLA domain (PF12041) was retrieved from the Pfam database (http://pfam.xfam.org). Then the pumpkin protein sequences were searched for this profile using hmmsearch tool of Tbtools software. *AtDELLA* genes were acquired by searching their gene IDs from TAIR (https://www.arabidopsis.org/), and then they were used as search queries to carry out BLASTp with the E-value of $1e^{-5}$ against pumpkin protein sequences. Subsequently, all the candidate protein sequences gained with the above two methods were submitted to CDD (http://ncbi.nlm.nih.gov/cdd) in NCBI to reconfirm the CmoDELLA proteins.

2.2 Chromosomal localization, physical and chemical properties of *CmoDELLA* genes

The chromosomal localization information of *CmoDELLA* family genes was obtained through cucurbit genomics database, and mapped using TBtools v 1.0986961 software (Chen et al., 2020). According to their chromosomal position, *CmoDELLA* family members were renamed. The physical and chemical properties, such as the coding sequence (CDS) lengths, amino acids number (AA), molecular weight (MW), isoelectric point (pI), grand average of hydropathicity (GRAVY) values and instability index, were predicted through Expasy (https//www.expasy.org/). Additionally, secondary structure prediction was executed *via* NPS@:SOPMA (https://npsa-prabi.Ibcp.fr/cgi-bin/npsa_automat.sopm.html).

2.3 Phylogenetic tree, gene structure and protein interaction networks analysis

DELLA protein sequences of pumpkin (7), cucumber (4), melon (4), watermelon (4), *Arabidopsis* (5), soybean (7), *Brassica napus* (13), rice (1), tomato (2) and maize (3) were retrieved from cucurbit genomics database, TAIR and Ensembl database (http://plants.ensembl.org/index.html), respectively. Based on multiple sequence alignment, phylogenetic tree was created using the neighbor-joining (NJ) method with MEGA 7.0. The conserved domains were analyzed by the NCBI CDD, and the conserved motifs were predicted by MEME (https://meme-suite.org/meme/). Furthermore, the distribution maps of conserved domains and conserved motifs were visualized using TBtools v 1.0986961. Promoter sequences (2 kp before the start codon) of *CmoDELLA* genes were analyzed through online PlantCare (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to obtain *cis*-regulatory elements. Interaction networks between CmoDELLA proteins and other proteins were conducted through online STRING (https://string-db.org/), using *Arabidopsis* as reference species.

2.4 Transcriptome sequencing analysis of *CmoDELLA* genes

The pre-published pumpkin transcriptome sequencing data under four tissues (roots, stems, leaves and fruits) (PRJNA385310) and NaCl stress (PRJNA437579) were obtained from the cucurbit genomics database for exploring the transcriptional profiles. Transcriptional levels were normalized by the reads per kilobase of exon per million reads mapped (RPKM) method. Expression heatmap of *CmoDELLA* genes was generated with TBtools v1.0986961.

2.5 Quantitative real-time PCR (qRT-PCR) and physiological indicators measurement under abiotic stress

In this current study, pumpkin variety "Hantailang" was used to explore the expression characterization of *CmoDELLA* genes under waterlogging stress and cold stress. "Hantailang" was cultivated and developed in a climate chamber with growth conditions (16 h light/8 h dark, 25°C daytime/16°C night). The pumpkin seedlings of two-leaf stage were exposed to abiotic stress. Waterlogging stress were conducted with water 2 cm above the soil surface. For cold stress, the seedlings were maintained at 15°C daytime/5°C night. The leaves were collected at day 10 after treatment and stored in -80 °C fridge for qRT-PCR analysis and physiological indicators measurement.

Total RNA of pumpkin leaves was extracted using TaKaRa MiniBEST Plant RNA Extraction Kit (Code No. 9769, TaKaRa, Dalian, China) with the instructions provided in the kit. First chain cDNA was synthesized using PrimeScript™ RT Master Mix (Perfect Real Time) Reagent Kit (Code No. RR036A, Takara, Dalian, China). Subsequently, qRT-PCR experiment was operated by TB Green® Premix Ex Taq™ II (Code No. RR820A, TaKaRa, Dalian, China) to display the expression patterns under waterlogging stress and cold stress. The primer sequences of *CmoDELLA* genes and reference gene (*CmoActin*) were listed in Table S1. All the experiments were carried out followed by the instructions. Each sample was replicated three times and the relative expression levels of *CmoDELLA* genes were analyzed with the $2^{-\Delta\Delta Ct}$ method.

The contents of proline and Malondialdehyde (MDA), catalase (CAT) activity in the leaves of pumpkin were measured using proline (Code No. BC0290, Solarbio, Beijing, China), MDA (Code No. BC0025, Solarbio, Beijing, China) and CAT (Code No. BC0205, Solarbio, Beijing, China) assay kit respectively, following the manufacturer's protocol. SOD activity was determined with SOD assay kit (Code No. G0104F, Geruisi, Suzhou, China).

3 Results

3.1 Identification and protein properties of *CmoDELLA* gene family

A total of 7 *CmoDELLA* genes were obtained through genome-wide search, and they were successively renamed as *CmoDELLA1~7* respectively, according to their chromosomal localization. Seven *CmoDELLA* genes were located on Chr1, Chr4, Chr11, Chr14 and Chr15, respectively. Among them, there were two *CmoDELLA* genes on the Chr4 (*CmoDELLA2* and *CmoDELLA3*) and Chr15 (*CmoDELLA6* and *CmoDELLA7*), while only one *CmoDELLA* gene on the Chr1, Chr11 and Chr14 (Figure 1). The protein properties of *CmoDELLA* gene family were collected in Table S2. The CDS lengths of *CmoDELLA* genes varied from 1605 bp to 1854 bp. AA number of CmoDELLA proteins ranged from 534 aa to 617 aa, the MW was between 58316.97 Da and 67383.21 Da, the pI range was 4.70 to 5.52, the GRAVY changed from -0.312 to -0.092. Seven CmoDELLA proteins were defined as unstable proteins with their instability index varying from 42.56 to 53.57. The secondary structures of seven CmoDELLA proteins were all made up α-helix, random coil, extended strand and β-sheets, in which α-helix accounted for the largest proportion (45.52%~49.17%), followed by random coil, extended strand and β-sheets, indicating that the structures of CmoDELLA proteins were the mixed type.

10.3389/fpls.2023.1137126

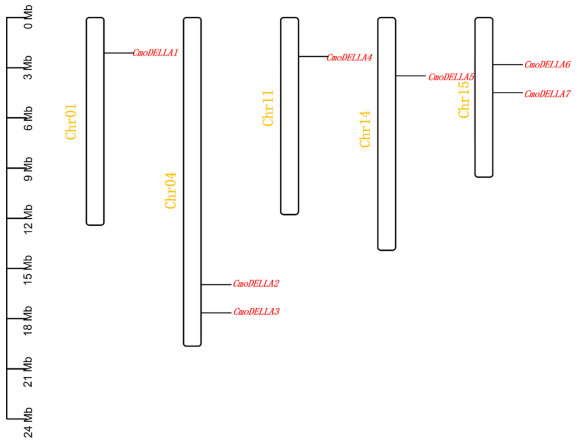


FIGURE 1
Chromosomal localization of *CmoDELLA* genes. The names of seven *CmoDELLA* genes are shown to the right of each chromosome. Gene position can be measured using the scale on the left of the figure in megabases (Mb).

Function annotation showed that *CmoDELLA1* and *CmoDELLA5* were related to the regulation of transcription, whereas *CmoDELLA2*, *CmoDELLA3*, *CmoDELLA4*, *CmoDELLA6* and *CmoDELLA7* were involved in the regulation of transcription, plant hormone signal transduction pathway, growth and development, and stress response, indicating their pivotal roles in regulating pumpkin development and stress response.

3.2 Phylogenetic tree, sequence alignment and structure of DELLA proteins

To evaluate the evolutionary relationship of DELLA proteins, phylogenetic tree was created by aligning 50 DELLA protein sequences from pumpkin, cucumber, melon, watermelon, *Arabidopsis*, soybean, *Brassica napus*, rice, tomato and maize (Figure 2). According to the phylogenetic relationship listed in the tree, the 50 DELLA proteins could be divided into 4 major classes: Class I, Class II, Class III and Class IV, which contained 8, 11, 15, and 16 members, respectively. As shown in the phylogenetic tree, DELLA proteins from pumpkin exhibited the relatively closer evolutionary relationship with those from watermeolon, melon, cucumber. In addition, CmoDELLA3 and CmoDELLA6 were classified into the Class II, CmoDELLA2 and CmoDELLA7 were clustered into Class III, CmoDELLA1, CmoDELLA4 and CmoDELLA5 belonged to Class IV. Multiple sequence alignment was performed among CmoDELLA proteins (Figure 3). The homology between CmoDELLA2 and CmoDELLA7 was the highest at 94.72%, followed by that between CmoDELLA3 and CmoDELLA6 at 89.37%. The homology of other CmoDELLA proteins was more than 64.66%, indicating the high conservation and complex evolution of CmoDELLA proteins. To better understand the structural differences of CmoDELLA proteins, the conserved domains and conserved motifs were detected. The results showed that CmoDELLA proteins were all composed of N-terminal DELLA domain and C-terminal GRAS domain (Figures 3, 4A), but the conserved motifs between CmoDELLA proteins were unevenly distributed (Figure 4B, Table 1). The number of motifs of different CmoDELLA proteins varied from 11 to 17. Class II memebers contained 17 motifs, Class III memebers included 16 to 17 motifs, whereas Class IV memebers held 11 to 14 motifs. A total of 20 motifs were found in CmoDELLA proteins, and 10 motifs (including motif1, motif2, motif3, motif5, motif6, motif7, motif8, motif9, motif10 and motif14) were highly conserved in all CmoDELLA proteins. Motif4 was identified in all CmoDELLA proteins except CmoDELLA4, and motif13 was present in all CmoDELLA proteins except CmoDELLA5. Moreover, motif15, motif16 and motif17 were detected in CmoDELLA2 and CmoDELLA7. Motif18, motif19 and motif20 were observed in CmoDELLA3 and CmoDELLA6. The different number of motifs between CmoDELLA proteins indicated their functional diversification in pumpkin.

3.3 *Cis*-regulatory elements of *CmoDELLA* genes

To learn more about the potential functions of *CmoDELLA* genes involved in different biological process, the promoter sequences of *CmoDELLA* genes were analyzed to identify *cis*-regulatory elements. The identified *cis*-regulatory elements were mainly related to light, plant hormone, stress, and plant growth and development (Figure 5). Among them, light response elements were the most abundant, such as G-Box, TCCC-motif, TCT-motif, Box II, I-box and AE-box. The plant hormone response elements were widely present in the promoter region, including abscisic acid response element (ABRE), TGACG-motif and CGTCA-motif for methyl jasmonic acid response element (MeJA), P-box and GARE-motif for gibberellin response element (GARE), TCA-element for salicylic acid response element (SARE) and AuxRR-core for auxin response element. Stress response elements containing MBS (involved in drought induction), LTR (involved in low temperature response), ARE (involved in anaerobic induction), STRE (involved in heat induction), WUN-motif (involved in wound response), and TC-rich repeats (involved in defense and stress response) were also identified. Simultaneously, we found some growth and development elements, for example, CAT-box associated with meristem, O2-site involved in the regulation of zein metabolism and circadian involved in circadian rhythm regulation.

3.4 Interaction networks of CmoDELLA proteins

To better understand the regulatory mechanism of CmoDELLA proteins, protein interaction networks were constructed. It could be seen from Figure 6 that seven CmoDELLA proteins, such as RGL1 (CmoDELLA3), GAI (CmoDELLA2, CmoDELLA4, CmoDELLA6 and CmoDELLA7), RGL2 (CmoDELLA1 and CmoDELLA5), all interacted with GID1A, GID1B, GID1C, SLY1 and PIF3 using *Arabidopsis* as reference species.

10.3389/fpls.2023.1137126

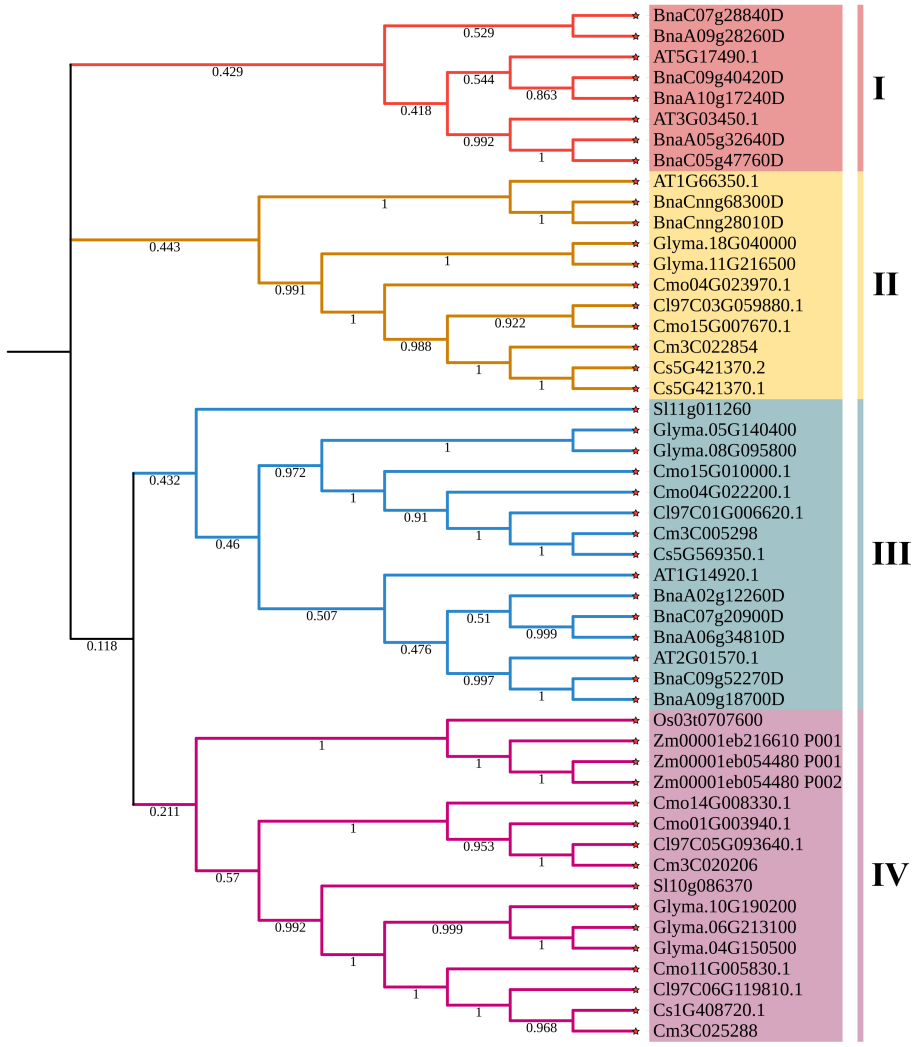


FIGURE 2
The phylogenetic tree of DELLA proteins from pumpkin, cucumber, melon, watermelon, *Arabidopsis*, soybean, *Brassica napus*, rice, tomato and maize. Cmo: pumpkin, Cs: cucumber, Cm: melon, Cl: watermelon, AT: *Arabidopsis*, Glyma.: soybean, Bna: *Brassica napus*, Os: rice, Sl: tomato, Zm: maize. The 4 major classes are represented by the different colors.

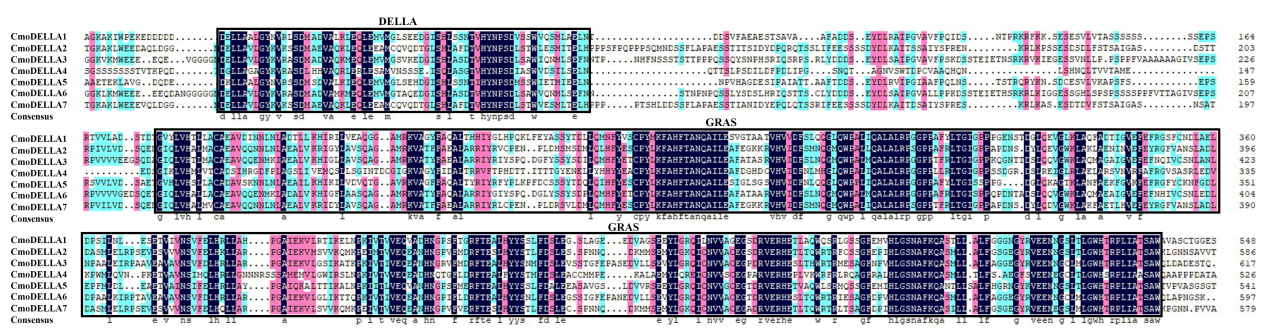


FIGURE 3
Amino acid sequence alignment of CmoDELLA proteins. The numbers on the right side of the sequence indicate the position of amino acid residues and the colors represent similarities in the protein sequences. The identical amino acid residues are shaded in black background, and the similar amino acid residues are shaded in the red and blue background.

10.3389/fpls.2023.1137126

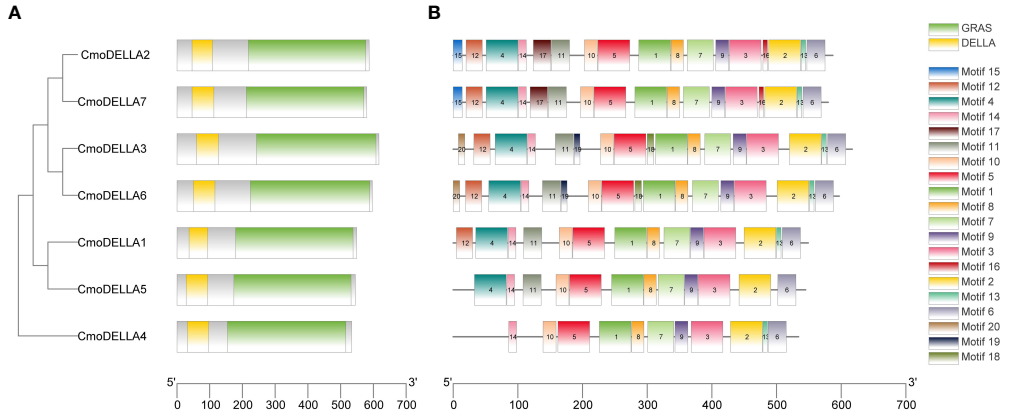


FIGURE 4
The conserved domains and conserved motifs of CmoDELLA proteins. The lengths of the conserved domains and conserved motifs are shown proportionally. **(A)** The conserved domains of CmoDELLA proteins. DELLA domain and GRAS domain are represented by yellow and green box, respectively. **(B)** The conserved motifs of CmoDELLA proteins. The different conserved motifs are indicated by different colored boxes.

3.5 Tissue-specific expression of *CmoDELLA* genes

To uncover the expression profiles of *CmoDELLA* genes, the transcriptional levels of *CmoDELLA* genes in the roots, stems, leaves and fruits were evaluated (Figure 7). RNA-seq data revealed that seven *CmoDELLA* genes were expressed in all analyzed tissues but exhibited different abundance levels. For example, *CmoDELLA2*, *CmoDELLA3*, *CmoDELLA4*, *CmoDELLA6* and *CmoDELLA7* had higher transcriptional profiles, whereas *CmoDELLA1* and *CmoDELLA5* demonstrated lower transcriptional levels in the four tissues. Moreover, *CmoDELLA2*, *CmoDELLA3* and

TABLE 1 The motif information in the CmoDELLA sequences.

Motif	Sequence	Number of Amino Acid
motif 1	QMHFYESCPYLKFAHFTANQAILEAFETAARVHVIDFSLNQGMQWPALIQ	50
motif 2	EMYLGRQICNVVACEGSDRVERHETLTQWRTRLESAGFEPIHLGSNAFKQ	50
motif 3	AIEKVLGVIKALKPKIVTVVEQEANHNGPVFMDRFTEALHYYSTLFDSLE	50
motif 4	LGYKVRSSDMADVAQKLEQLEMVMGQVZEDGISHLASDTVHYNPSDLSSW	50
motif 5	SLVHALFACAEAVRVENNNLAEALGKHIRPLIATQAGAMRK	41
motif 6	DSLQEVGWKLAQFAETIGVEFEFRGFVCNNLADLDPSMLELRPEEVEAV	49
motif 7	ALALRPGGPPAFRLTGIGPP	20
motif 8	VATYFAZALARRIYRJYPPKP	21
motif 9	AEYSDDSEYDLKAIPGVAIFPPKDSSTEK	29
motif 10	VFELHRLLARP	11
motif 11	KGZCSSLSGGKAKLWEEEEQEDGGGD	26
motif 12	AGASSEPSRPVVLVDSQETG	20
motif 13	VZSMLSELNNPPS	13
motif 14	EGFRVEENEGCLMLGWHSRPLIAASAWK	28
motif 15	PNNQDKMM	8
motif 16	DSSFLAPAESSTIANIDYEPQRQTSSRI	28
motif 17	MKMKRE	6
motif 18	ETNSRKRLKI	10
motif 19	FECASSYTD	9
motif 20	PQSSQYSDPHHRIQ	14

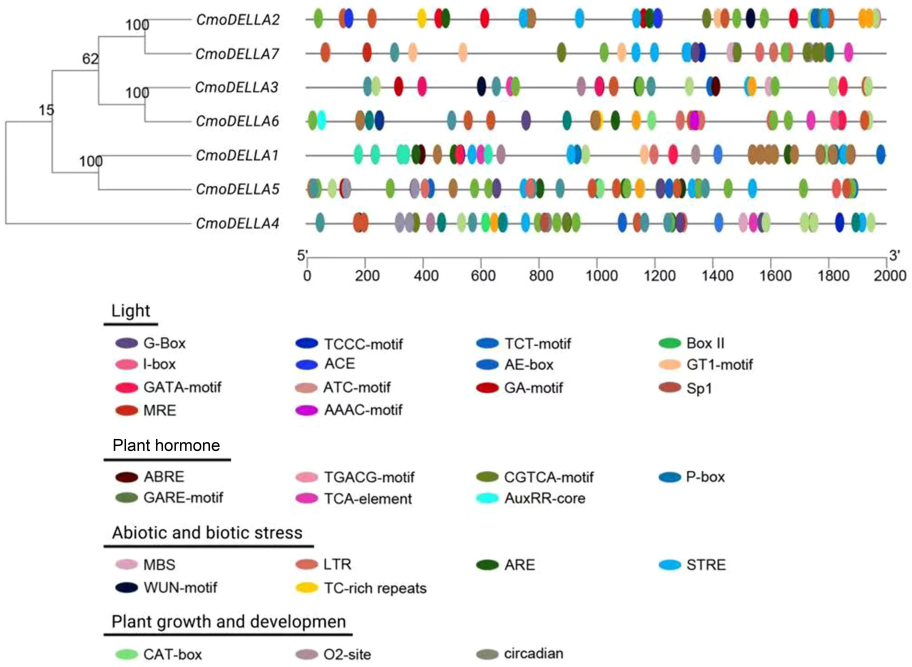


FIGURE 5
Cis-regulatory elements in the promoter regions of *CmoDELLA* genes. These *cis*-regulatory elements are related to light response, plant hormone response, stress response, and plant growth and development. The different colored circles represent the different types and positions of *cis*-regulatory elements in each *CmoDELLA* gene.

CmoDELLA7 were highly expressed in the stems, while *CmoDELLA1*, *CmoDELLA4*, *CmoDELLA5* and *CmoDELLA6* were mainly expressed in the fruits. Differential expression of *CmoDELLA* genes in different tissues indicated that the tissue specificity and functional divergence of *CmoDELLA* genes in the growth and development of pumpkin.

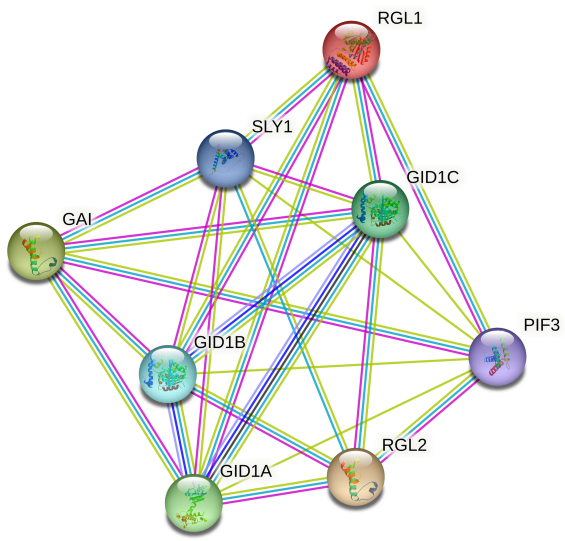


FIGURE 6
Interaction networks of CmoDELLA proteins. The network nodes represent proteins and the edges indicate protein-protein associations.

3.6 Transcriptome and qRT-PCR analysis of *CmoDELLA* genes under abiotic stress

CmoDELLA genes not only participated in plant development, but were also involved in various abiotic stress. Promoter *cis*-regulatory elements analysis also showed that *CmoDELLA* genes contained some stress response elements. To further explore and gain more insights into possible functions under abiotic stresses, such as salt, waterlogging and cold temperature, transcriptional profiles of *CmoDELLA* genes under NaCl stress were analyzed based on the pre-published RNA-seq data. *CmoDELLA* genes showed significant differences in response to NaCl stress (Figure 8). Transcriptional levels of *CmoDELLA2*, *CmoDELLA3*, *CmoDELLA4*, *CmoDELLA6* and *CmoDELLA7* were down-regulated after 75 mmol/L NaCl stress for 24 h in contrast with the normal condition. In contrast, the expression levels of *CmoDELLA1* and *CmoDELLA5* showed an increased trend after NaCl stress. In addition, the expression levels of *CmoDELLA* genes were investigated by qRT-PCR for further understanding the functions in response to waterlogging stress and cold stress. As shown in Figure 9, seven *CmoDELLA* genes were all up-regulated under waterlogging stress and cold stress after 10 days, compared to the normal condition. Under waterlogging stress, the expression levels of *CmoDELLA* genes were significantly higher than that without waterlogging treatment, except for *CmoDELLA4*. For example, *CmoDELLA1~7* showed 1.62-, 1.51-, 1.45-, 1.18-, 1.30-, 1.44-, 1.41-fold higher expression levels, respectively, under waterlogging stress. Under cold stress, *CmoDELLA1~7* exhited 9.25-, 8.56-, 7.22-, 5.81-, 6.38-, 6.95-, 5.02-fold remarkably higher expression, respectively, compared to the control. The above results suggested that

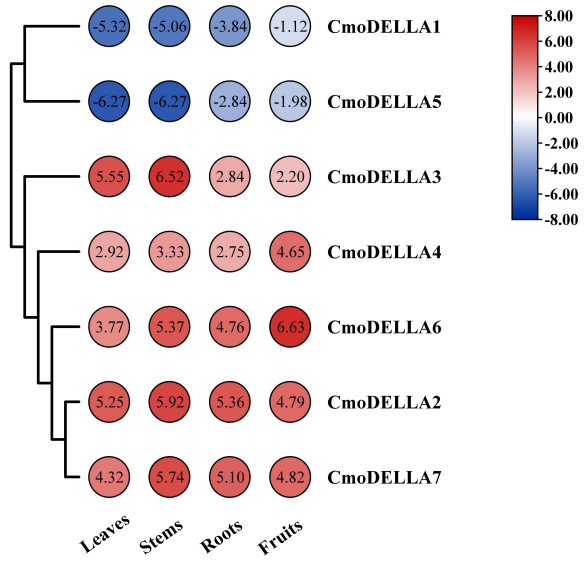


FIGURE 7
Transcriptional levels of *CmoDELLA* genes in the different tissues (leaves, stems, roots and fruits). The transcriptional levels are normalized by the reads per kilobase of exon per million reads mapped (RPKM) method. The red color represents the higher expression, while the blue color indicates the lower expression.

CmoDELLA genes might be implicated in response to NaCl stress, waterlogging stress and cold stress in pumpkin, however, the functions of different *CmoDELLA* genes varied under different stress.

3.7 Physiological changes under waterlogging stress and cold stress

In this study, we measured the changes of proline, MDA, SOD and CAT in pumpkin leaves under waterlogging stress and cold stress after 10 days (Figure 10). Under waterlogging stress and cold stress, the proline contents were remarkably lower, compared to the control. The MDA contents were 1.44- and 1.62-fold of that in the control, respectively, but did not differ significantly between the control and waterlogging stress. The CAT activities were significantly decreased by 18.16% and 35.43% under waterlogging stress and cold stress, respectively, compared with that in the control. While, SOD activity was lower under waterlogging stress than that in the control, but significantly higher under cold stress than that in the control.

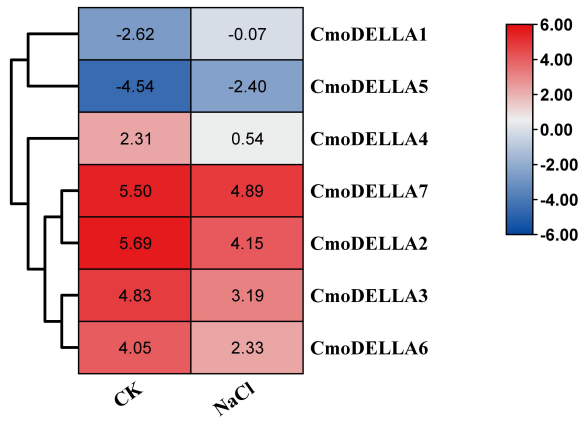


FIGURE 8
Transcriptional profiles of *CmoDELLA* genes under NaCl stress after 24 h. The transcriptional levels are normalized by RPKM method. The red color represents the higher expression, while the blue color indicates the lower expression.

4 Discussion

DELLA gene family plays key roles in regulating plant development and stress response. Up to now, *DELLA* genes have been extensively identified and characterized in plants, but the research of *DELLA* genes in pumpkin has rarely been reported. In this present study, seven *CmoDELLA* genes were obtained in pumpkin by genome-wide analysis, which were located on five chromosomes respectively. CmoDELLA proteins were all unstable proteins, and the secondary structures of them were mainly made up α-helix and random coil. The phylogenetic tree displayed that 50 DELLA protein sequences from pumpkin, cucumber, melon, watermelon, *Arabidopsis*, soybean, *Brassica napus*, rice, tomato and maize were divided into four subfamilies, and the DELLA proteins of pumpkin shared the closer evolutionary relationship with those of watermeolon, melon, cucumber, which may be due to pumpkin and watermeolon, melon, cucumber belonging to cucurbita family. N-terminal DELLA domain and C-terminal GRAS domain were the typical conserved domains of DELLA family members in various plants (Sarwar et al., 2021). In this study, it was found that seven CmoDELLA proteins held the highly conserved DELLA domain and GRAS domain, shared 10 conserved motifs, however, motif numbers between CmoDELLA proteins were unevenly distributed, indicating the complex evolution and functional diversification of CmoDELLA proteins. These findings were consistent with the study has been found in BnaDELLA proteins (Sarwar et al., 2021).

Promoter *cis*-regulatory elements determine the specific function of the genes. Analysis of *cis*-regulatory elements can provide an insight into exploring the expression and regulation mechanism of genes under different tissues and stress environments. Previous studies have shown that DELLA proteins participate in plant hormone signal transduction pathway, including GA, auxin, abscisic acid, ethylene, and jasmonate (Achard et al., 2006), which affects diverse aspects of plant development and response to environmental stress (Xu et al., 2014). Additionally, some researches have reported the importance of ABRE, SARE and MeJA for abiotic stress tolerance *via* plant hormone signal transduction pathway (Doornbos et al., 2011; Rivas-San Vicente and Plasencia, 2011). In the current study, *CmoDELLA* genes contained a lot of promoter *cis*-regulatory elements involved in light, plant hormone, stress, and plant growth and development, such as ABRE, MeJA, GARE, SARE and ARE, suggesting that the *CmoDELLA* genes may be responsible for plant development and stress response in

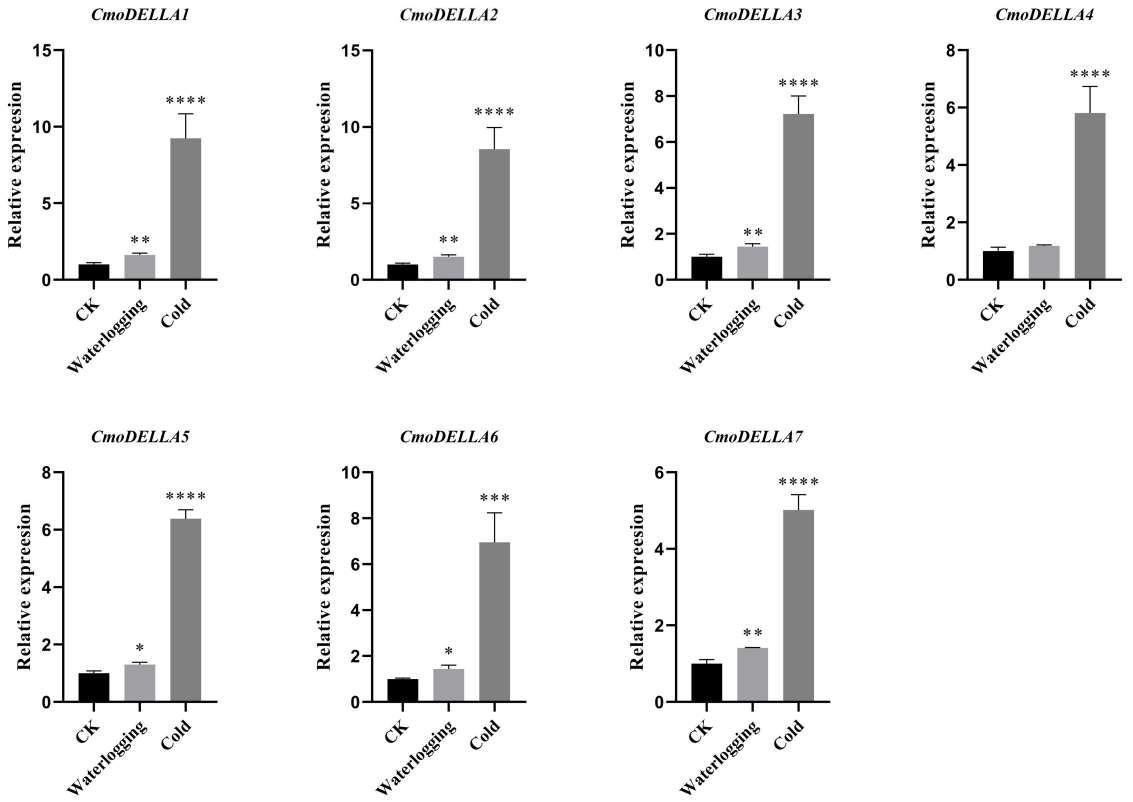


FIGURE 9
Expression levels of *CmoDELLA* genes in response to waterlogging stress and cold stress after 10 days. The relative expression levels of *CmoDELLA* genes are normalized with respect to the reference gene (*CmoActin*). The X-axis represents the waterlogging stress and cold stress. The Y-axis represents the relative expression levels. The values are denoted as the means ± SDs. The significant difference is represented by asterisks at *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

pumpkin by plant hormone signal pathway. Similar findings have been also reported in apple (Fan et al., 2017), rice (Muhammad et al., 2019), soybean (Liang et al., 2022) and mango (Wang et al., 2022).

Usually, DELLA proteins play essential roles by interacting with other TFs in plants, as key negative regulators of GA signaling (Zhou et al., 2020). GA signal transduction pathway is mainly involved in GID1, DELLA proteins and F-box family proteins (Xue et al., 2022). GID1A, GID1B and GID1C are soluble GA receptors, and SLY1 belongs to F-box family protein. GID1 interacts with the N-terminal domain of DELLA proteins to form GID1-DELLA complex, and then DELLA proteins are bound by SLY1, which leading to degradation of DELLA proteins (McGinnis et al., 2003; Murase et al., 2008). PIF3 is phytochrome-associated protein, and DELLA proteins bind with the promoter of PIF3 to inhibit the PIF-mediated hypocotyl elongation (de Lucas et al., 2008). Protein interaction networks analysis indicated that seven DELLA proteins all interacted with GID1A, GID1B, GID1C, SLY1 and PIF3, suggested their involvement in the development process *via* GA signal transduction pathway.

Gene functions are usually understood by detecting tissue-specific expression profiles of genes. Here, the transcriptome data of *CmoDELLA* genes was analyzed in the roots, stems, leaves and fruits. The results showed that seven genes were all expressed in the four tissues. Moreover, *CmoDELLA2*, *CmoDELLA3* and *CmoDELLA7* were highly expressed in the stems, while *CmoDELLA1*, *CmoDELLA4*, *CmoDELLA5* and *CmoDELLA6* were mainly expressed in the fruits. In cucumber, four *DELLA* genes displayed the distinct expression patterns in the different tissues, and *CsGAIP* exhibited the higher expression levels in the stems (Zhang et al., 2014). In compliance with this, some researches have also been reported that *DELLA* genes have a predominant roles in regulating plant stem elongation growth in *Arabidopsis* (King et al., 2001) and *Brassica napus* (Zhao et al., 2017). Furthermore, we found that *CmoDELLA1* showed significantly up-regulated expression in the pollinated fruits than that in the ovaries without pollination in pumpkin (Luo et al., 2021). Recent study showed that DELLA proteins promoted ovule initiation by interacting with the CUC2 TF (Barro-Trastoy et al., 2022). The results presented here suggested that *CmoDELLA2*, *CmoDELLA3* and *CmoDELLA7* were related to the stems development, *CmoDELLA1*, *CmoDELLA4*, *CmoDELLA5* and *CmoDELLA6* were associated with the fruits development. Further investigation should be conducted to verify their functions.

Several studies have reported that *DELLA* genes participate in the stress response process (Huang et al., 2006). For instance, overaccumulation of DELLA proteins enhances the salt stress (Sakuraba et al., 2017) and cold stress tolerance (Yang et al.,

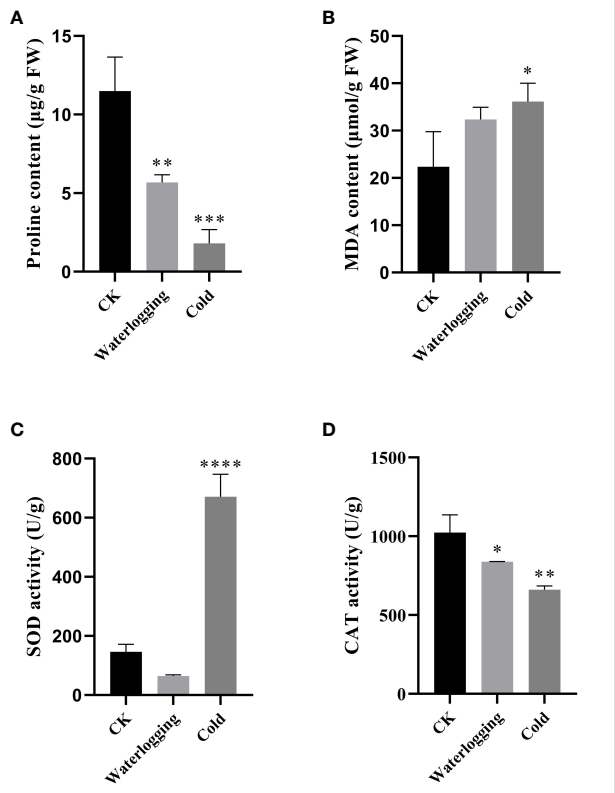


FIGURE 10
Physiological changes under waterlogging stress and cold stress after 10 days in pumpkin. **(A)**; proline content. **(B)**; MDA content. **(C)**; SOD activity. **(D)**; CAT activity. The values are denoted as the means ± SDs. The significant difference is represented by asterisks at *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

2013), which significantly improves plant fitness. Han et al. (2020) analyzed the transcriptome data of *DELLA* genes in *Camellia sinensis* under drought stress, cold stress and NaCl stress, and speculated that *DELLA* genes were involved in the abiotic stress response. Sarwar et al. (2021) found that *BnaDELLA* genes exhibited different expression abundance under NaCl stress, suggesting the *BnaDELLA* genes vital roles in susceptibility to NaCl stress. Consistent with those, this current study showed the distinct transcriptional patterns of *CmoDELLA* genes under NaCl stress. For instance, *CmoDELLA1* and *CmoDELLA5* showed an increased expression under NaCl stress after 24 h, indicating *CmoDELLA1* and *CmoDELLA5* might be crucial in response to NaCl stress. Additionally, in our qRT-PCR analysis, *CmoDELLA1*, *CmoDELLA2*, *CmoDELLA3*, *CmoDELLA5*, *CmoDELLA6* and *CmoDELLA7* were remarkably induced under waterlogging stress after 10 days. And all of the 7 *CmoDELLA* genes showed significantly induced expression under cold stress after 10 days. These results suggested the vital roles of *CmoDELLA* genes in response to waterlogging stress and cold stress. Many previous studies on *AtDELLA* genes (Zhou et al., 2017; Blanco-Touriñán et al., 2020) and *BnaDELLA* genes (Sarwar et al., 2021) have provided evidence of their fundamental roles in regulating plant physiology under cold stress. Moreover, the expression profiles of different *BnaDELLA* genes varied under different stress treatments (Sarwar et al., 2021). Based on the above results, *CmoDELLA* genes might mediate the stress response of pumpkin to NaCl, waterlogging and cold, however, the functions of different *CmoDELLA* genes varied under different stress.

Previous report has shown that adverse environment conditions promote DELLA accumulation (Achard et al., 2008b) and then DELLA proteins increase SOD activity to improve salt stress tolerance in wheat (Wang et al., 2016). We here showed that *CmoDELLA* genes expression and SOD activity were significantly higher under cold stress than those in the control. Further researches are required to better comprehend the regulatory relationships between *CmoDELLA* genes expression and SOD activity under abiotic stress.

5 Conclusions

In summary, seven *CmoDELLA* genes were obtained in pumpkin by genome-wide analysis. Furthermore, the chromosomal localization, protein properties, phylogenetic tree, gene structure, promoter *cis*-regulatory elements and protein interaction networks of *CmoDELLA* genes were conducted. Expression profiles of *CmoDELLA* genes under different tissues and abiotic stress were determined through RNA-seq data and qRT-PCR. Additionally, physiological changes of pumpkin seedlings were measured under waterlogging stress and cold stress. As a whole, these results revealed the vital roles of *CmoDELLA* genes in regulating plant development and stress response in pumpkin, which would provide valuable clues for further studying the potential functions and regulatory networks of *CmoDELLA* genes.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: NCBI PRJNA385310 and PRJNA437579.

Author contributions

YS conceived and designed the experiments. WL carried out the experiments. WL, ZZ, HC, WA, LL, and JL analyzed the data, prepared figures and tables. WL wrote the manuscript. YS and XL reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Key Research and Development Program of Henan Province (No. 212102110130), the Major Public Welfare Projects of Henan Province (No. 201300111300) and the Key Science and Technology Program of Xinxiang City (No. GG2019013).

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.2023.1137126/full#supplementary-material

References

Achard, P., Cheng, H., De Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., et al. (2006). Integration of plant responses to environmentally activated phytohormonal signals. *Science* 311, 91–94. doi: 10.1126/science.1118642

Achard, P., Gong, F., Cheminant, S., Alioua, M., Hedden, P., and Genschik, P. (2008a). The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins *via* its effect on gibberellin metabolism. *Plant Cell* 20, 2117–2129. doi: 10.1105/tpc.108.058941

Achard, P., Liao, L. L., Jiang, C. F., Desnos, T., Bartlett, J., Fu, X. D., et al. (2007). DELLAs contribute to plant photomorphogenesis. *Plant Physiol.* 143, 1163–1172. doi: 10.1104/pp.106.092254

Achard, P., Renou, J. P., Berthomé, R., Harberd, N. P., and Genschik, P. (2008b). Plant DELLAs restrain growth and promote survival of adversity by reducing the levels of reactive oxygen species. *Curr. Biol.* 18, 656–660. doi: 10.1016/j.cub.2008.04.034

Achard, P., Vriezen, W. H., van der Straeten, D., and Harberd, N. P. (2003). Ethylene regulates arabidopsis development via the modulation of DELLA protein growth repressor function. *Plant Cell* 15, 2816–2825. doi: 10.1105/tpc.015685

Barro-Trastoy, D., Gomez, M. D., Blanco-Touriñán, N., Tornero, P., and Perez-Amador, M. A. (2022). Gibberellins regulate ovule number through a DELLA-CUC2 complex in *Arabidopsis*. *Plant J.* 110, 43–57. doi: 10.1111/tpj.15607

Blanco-Touriñán, N., Legris, M., Minguet, E. G., Costigliolo-Rojas, C., Nohales, M. A., Iniesto, E., et al. (2020). COP1 destabilizes DELLA proteins in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 117, 13792–13799. doi: 10.1073/pnas.1907969117

Cao, D. N., Cheng, H., Wu, W., Soo, H. M., and Peng, J. R. (2006). Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in *Arabidopsis*. *Plant Physiol.* 142, 509–525. doi: 10.1104/pp.106.082289

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

Chen, J. H., Cheng, T. L., Wang, P. K., Tian, L., Wang, G. P., Luo, Y. M., et al. (2013). Genome-wide bioinformatics analysis of DELLA-family proteins from plants. *Plant Omics* 6, 201–207.

Davière, J. M., Wild, M., Regnault, T., Baumberger, N., Eisler, H., Genschik, P., et al. (2014). Class I TCP-DELLA interactions in inflorescence shoot apex determine plant height. *Curr. Biol.* 24, 1923–1928. doi: 10.1016/j.cub.2014.07.012

de Lucas, M., Davière, J. M., Rodríguez-Falcón, M., Pontin, M., Iglesias-Pedraz, J. M., Lorrain, S., et al. (2008). A molecular framework for light and gibberellin control of cell elongation. *Nature* 451, 480–484. doi: 10.1038/nature06520

Doornbos, R. F., Geraats, B. P., Kuramae, E. E., Van Loon, L. C., and Bakker, P. A. (2011). Effects of jasmonic acid, ethylene, and salicylic acid signaling on the rhizosphere bacterial community of arabidopsis thaliana. *Mol. Plant Microbe Interact.* 24, 395–407. doi: 10.1094/mpmi-05-10-0115

Fan, S., Zhang, D., Zhang, L. Z., Gao, C., Xin, M. Z., Tahir, M. M., et al. (2017). Comprehensive analysis of GASA family members in the malus domestica genome: identification, characterization, and their expressions in response to apple flower induction. *BMC Genomics* 18, 827. doi: 10.1186/s12864-017-4213-5

Feng, S. H., Martinez, C., Gusmaroli, G., Wang, Y., Zhou, J. L., Wang, F., et al. (2008). Coordinated regulation of *Arabidopsis thaliana* development by light and gibberellins. *Nature* 451, 475–479. doi: 10.1038/nature06448

Guan, H. L., Huang, X. M., Zhu, Y. N., Xie, B. X., Liu, H. C., Song, S. W., et al. (2021). Identification of DELLA genes and key stage for GA sensitivity in bolting and flowering of flowering Chinese cabbage. *Int. J. Mol. Sci.* 22, 12092. doi: 10.3390/ijms222212092

Han, Y. X., Dai, H. W., Zheng, S. T., Tong, H. R., and Yuan, L. Y. (2020). Identification and expression analysis of the DELLA gene family in camellia sinensis (L.) o. ktze. *Plant Sci. J.* 38, 644–653. doi: 10.11913/PSJ.2095-0836.2020.50644

Hu, J. H., Israeli, A., Ori, N., and Sun, T. P. (2018). The interaction between DELLA and ARF/IAA mediates crosstalk between gibberellin and auxin signaling to control fruit initiation in tomato. *Plant Cell* 30, 1710–1728. doi: 10.1105/tpc.18.00363

Huang, X. Z., Jiang, C. F., Liao, L. L., and Fu, X. D. (2006). Progress on molecular foundation of GA biosynthesis pathway and signaling. *Chin. Bull. Bot.* 5, 499–510.

King, K. E., Moritz, T., and Harberd, N. P. (2001). Gibberellins are not required for normal stem growth in arabidopsis thaliana in the absence of GAI and RGA. *Genetics* 159, 767–776. doi: 10.1093/genetics/159.2.767

Li, K. L., Yu, R. B., Fan, L. M., Wei, N., Chen, H. D., and Deng, X. W. (2016). DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in *Arabidopsis*. *Nat. Commun.* 7, 11868. doi: 10.1038/ncomms11868

Li, L., Shu, S., Xu, Q., An, Y. H., Sun, J., and Guo, S. R. (2017). NO accumulation alleviates H_2O_2-dependent oxidative damage induced by $Ca(NO_3)_2$ stress in the leaves of pumpkin-grafted cucumber seedlings. *Physiol. Plant* 160, 33–45. doi: 10.1111/ppl.12535

Li, Q. F., Wang, C. M., Jiang, L., Li, S., Sun, S. S., and He, J. X. (2012). An interaction between BZR1 and DELLAs mediates direct signaling crosstalk between brassinosteroids and gibberellins in *Arabidopsis*. *Sci. Signal* 5, ra72. doi: 10.1126/scisignal.2002908

Li, W. J., Zhang, J. X., Sun, H. Y., Wang, S. M., Chen, K. Q., Liu, Y. X., et al. (2018). FveRGA1, encoding a DELLA protein, negatively regulates runner production in fragaria vesca. *Planta* 247, 941–951. doi: 10.1007/s00425-017-2839-9

Li, X. L., Liu, W., Li, B., Liu, G. Y., Wei, Y. X., He, C. Z., et al. (2018). Identification and functional analysis of cassava DELLA proteins in plant disease resistance against cassava bacterial blight. *Plant Physiol. Biochem.* 124, 70–76. doi: 10.1016/j.plaphy.2017.12.022

Liang, N. S., Zhan, Y. G., Yu, L., Wang, Z. Q., and Zeng, F. S. (2019). Characteristics and expression analysis of FmTCP15 under abiotic stresses and hormones and interact with DELLA protein in fraxinus mandshurica rupr. *Forests* 10, 343–343. doi: 10.3390/f10040343

Liang, S., Chen, Q. S., Zhu, Z. K., Li, D. D., Qi, Z. M., and Xin, D. W. (2022). Identification and analysis of soybean DELLA gene family. *Chin. J. Oil Crop Sci.* 44, 996–1005. doi: 10.19802/j.issn.1007-9084.2021224

Liao, Z. G., Yu, H., Duan, J. B., Yuan, K., Yu, C. J., Meng, X. B., et al. (2019). SLR1 inhibits MOC1 degradation to coordinate tiller number and plant height in rice. *Nat. Commun.* 10, 2738. doi: 10.1038/s41467-019-10667-2

Lim, S., Park, J., Lee, N., Jeong, J., Toh, S., Watanabe, A., et al. (2013). ABA-insensitive3, ABA-insensitive5, and DELLAs interact to activate the expression of SOMNUS and other high-temperature-inducible genes in imbibed seeds in *Arabidopsis*. *Plant Cell* 25, 4863–4878. doi: 10.1105/tpc.113.118604

Liu, H., Yang, H. Z., Li, L., Wang, F., Wu, S. H., and Li, Y. Y. (2016). Cloning and expression analysis of *PpGAI* genes of DELLA protein related to dormancy from pyrus. *Mol. Plant Breed* 14, 1995–2002. doi: 10.13271/j.mpb.014.001995

Liu, K., Li, Y. H., Chen, X. N., Li, L. J., Liu, K., Zhao, H. P., et al. (2018). ERF72 interacts with ARF6 and BZR1 to regulate hypocotyl elongation in *Arabidopsis*. *J. Exp. Bot.* 69, 3933–3947. doi: 10.1093/jxb/ery220

Liu, S. S., Li, H., Lv, X. Z., Ahammed, G. J., Xia, X. J., Zhou, J., et al. (2016). Grafting cucumber onto luffa improves drought tolerance by increasing ABA biosynthesis and sensitivity. *Sci. Rep.* 6, 20212. doi: 10.1038/srep20212

Luo, W. R., Li, Y. Y., Sun, Y. D., Lu, L., Zhao, Z. X., Zhou, J. G., et al. (2021). Comparative RNA-seq analysis reveals candidate genes associated with fruit set in pumpkin. *Sci. Hortic.* 288, 110255. doi: 10.1016/j.scienta.2021.110255

Martí, C., Orzáez, D., Ellul, P., Moreno, V., Carbonell, J., and Granell, A. (2007). Silencing of DELLA induces facultative parthenocarpy in tomato fruits. *Plant J.* 52, 865–876. doi: 10.1111/j.1365-313X.2007.03282.x

McGinnis, K. M., Thomas, S. G., Soule, J. D., Strader, L. C., Zale, J. M., Sun, T. P., et al. (2003). The *Arabidopsis SLEEPY1* gene encodes a putative f-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell* 15, 1120–1130. doi: 10.1105/tpc.010827

Muhammad, I., Li, W. Q., Jing, X. Q., Zhou, M. R., Shalmani, A., Ali, M., et al. (2019). A systematic in silico prediction of gibberellic acid stimulated GASA family members: a novel small peptide contributes to floral architecture and transcriptomic changes induced by external stimuli in rice. *J. Plant Physiol.* 234–235, 117-132. doi: 10.1016/j.jplph.2019.02.005

Murase, K., Hirano, Y., Sun, T. P., and Hakoshima, T. (2008). Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* 456, 459–463. doi: 10.1038/nature07519

Navarro, L., Bari, R., Achard, P., Lisón, P., Nemri, A., Harberd, N. P., et al. (2008). DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. *Curr. Biol.* 18, 650–655. doi: 10.1016/j.cub.2008.03.060

Oh, E., Zhu, J. Y., Bai, M. Y., Arenhart, R. A., Sun, Y., and Wang, Z. Y. (2014). Cell elongation is regulated through a central circuit of interacting transcription factors in the arabidopsis hypocotyl. *Elife* 3, e03031. doi: 10.7554/eLife.03031

Rivas-San Vicente, M., and Plasencia, J. (2011). Salicylic acid beyond defence: its role in plant growth and development. *J. Exp. Bot.* 62, 3321–3338. doi: 10.1093/jxb/err031

Sakuraba, Y., Bülbül, S., Piao, W. L., Choi, G., and Paek, N. C. (2017). *Arabidopsis* EARLY FLOWERING3 increases salt tolerance by suppressing salt stress response pathways. *Plant J.* 92, 1106–1120. doi: 10.1111/tpj.13747

Sarwar, R., Jiang, T., Ding, P., Gao, Y., Tan, X. L., and Zhu, K. M. (2021). Genome-wide analysis and functional characterization of the DELLA gene family associated with stress tolerance in b. napus. *BMC Plant Biol.* 21, 286. doi: 10.1186/s12870-021-03054-x

Sechet, J., Frey, A., Effroy-Cuzzi, D., Berger, A., Perreau, F., Cueff, G., et al. (2016). Xyloglucan metabolism differentially impacts the cell wall characteristics of the endosperm and embryo during *Arabidopsis* seed germination. *Plant Physiol.* 170, 1367–1380. doi: 10.1104/pp.15.01312

Thines, B., Katsir, L., Melotto, M., Niu, Y. J., Mandaokar, A., Liu, G. H., et al. (2007). JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* 448, 661–665. doi: 10.1038/nature05960

Wang, J. W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, R. Q., Fan, X. C., Song, M. F., Xiao, Y., Guo, L., Meng, F. H., et al. (2016). A wheat DELLA gain-of-function mutant aibian 1 promotes seedlings salt tolerance. *Acta Agronomica Sin.* 42, 1721–1726. doi: 10.3724/SP.J.1006.2016.01721

Wang, Y., He, S., Wei, Y. Z., Dong, C., Liu, L. Q., Jue, D. W., et al. (2020). Molecular and functional characterization of two DELLA protein-coding genes in litchi. *Gene* 738, 144455. doi: 10.1016/j.gene.2020.144455

Wang, Y. Y., Yu, H. P., Zhao, Z. C., Gao, A. P., Zhang, Y., and Zhou, K. B. (2022). Cloning and analysis of. *DELLA-GAI Gene its promoter mango. Mol. Plant Breed* 20, 2526–2533. doi: 10.13271/j.mpb.020.002526

Worldmapper (2021). Available at: http://worldmapper.org/maps/pumpkinproduction/ (Accessed December 2, 2022).

Wu, G., Park, M. Y., Conway, S. R., Wang, J. W., Weigel, D., and Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138, 750–759. doi: 10.1016/j.cell.2009.06.031

Xu, H., Liu, Q., Yao, T., and Fu, X. D. (2014). Shedding light on integrative GA signaling. *Curr. Opin. Plant Biol.* 21, 89–95. doi: 10.1016/j.pbi.2014.06.010

Xue, H. D., Gao, X., He, P., and Xiao, G. H. (2022). Origin, evolution, and molecular function of DELLA proteins in plants. *Crop J.* 10, 287–299. doi: 10.1016/j.cj.2021.06.005

Yang, D. L., Dong, W. X., Zhang, Y. Y., and He, Z. H. (2013). Gibberellins modulate abiotic stress tolerance in plants. *Sci. Sin.* 43, 1119–1126. doi: 10.1360/052013-321

Zhang, J., Yang, J. J., Yang, Y., Luo, J., Zheng, X. Y., Wen, C. L., et al. (2019). Transcription factor *CsWIN1* regulates pericarp wax biosynthesis in cucumber grafted on pumpkin. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01564

Zhang, Y., Liu, B., Yang, S., An, J. B., Chen, C. H., Zhang, X. L., et al. (2014). A cucumber *DELLA* homolog *CsGAIP* may inhibit staminate development through transcriptional repression of b class floral homeotic genes. *PloS One* 9, e91804. doi: 10.1371/journal.pone.0091804

Zhao, B., Li, H. T., Li, J. J., Wang, B., Cheng, D., Wang, J., et al. (2017). Brassica napus DS-3, encoding a DELLA protein, negatively regulates stem elongation through gibberellin signaling pathway. *Theor. Appl. Genet.* 130, 727–741. doi: 10.1007/s00122-016-2846-4

Zhou, M. Q., Chen, H., Wei, D. H., Ma, H., and Lin, J. (2017). *Arabidopsis CBF3* and *DELLAs* positively regulate each other in response to low temperature. *Sci. Rep.* 7, 39819. doi: 10.1038/srep39819

Zhou, P., Li, Q. F., Xiong, M., Fan, X. L., Zhao, D. S., Zhang, C. Q., et al. (2020). Advances in DELLA protein-mediated phytohormonal crosstalk in regulation of plant growth and development. *Plant Physiol. J.* 56, 661–671. doi: 10.13592/j.cnki.ppj.2019.0570

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/fpls.2023.1144317

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Biao Jiang,
Guangdong Academy of Agricultural Sciences, China
Guangwei Zhao,
Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Peng Gao
gaopeng_neau@163.com
Sikandar Amanullah
sikandaraman@yahoo.com

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

RECEIVED 14 January 2023
ACCEPTED 06 February 2023
PUBLISHED 24 February 2023

CITATION
Zhang T, Xu N, Amanullah S and Gao P (2023) Genome-wide identification, evolution, and expression analysis of *MLO* gene family in melon (*Cucumis melo* L.). *Front. Plant Sci.* 14:1144317. doi: 10.3389/fpls.2023.1144317

Genome-wide identification, evolution, and expression analysis of *MLO* gene family in melon (*Cucumis melo* L.)

Taifeng Zhang[1,2], Nan Xu[1,2], Sikandar Amanullah[1,2]* and Peng Gao[1,2]*

[1]Key Laboratory of Biology and Genetic Improvement of Horticulture Crops (Northeast Region), Ministry of Agriculture and Rural Affairs, Harbin, China, [2]College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, China

Powdery mildew (PM) is one of the main fungal diseases that appear during the cultivation of the melon fruit crop. Mildew Resistance Locus "O" (MLO) is known as a gene family and has seven conserved transmembrane domains. An induced functional loss of a specific *MLO* gene could mainly confer PM resistance to melons. However, the genomic structure of *MLO* genes and its main role in PM resistance still remain unclear in melon. In this study, bioinformatic analysis identified a total of 14 *MLO* gene family members in the melon genome sequence, and these genes were distributed in an uneven manner on eight chromosomes. The phylogenetic analysis divided the *CmMLO* genes into five different clades, and gene structural analysis showed that genes in the same clade had similar intron and exon distribution patterns. In addition, by cloning the *CmMLO* gene sequence in four melon lines, analyzing the *CmMLO* gene expression pattern after infection, and making microscopic observations of the infection pattern of PM, we concluded that the *CmMLO5* (*MELO3C012438*) gene plays a negative role in regulating PM-resistance in the susceptible melon line (Topmark), and the critical time point for gene function was noticed at 24 and 72 hours after PM infection. The mutational analysis exhibited a single base mutation at 572 bp, which further results in loss of protein function, thus conferring PM resistance in melon. In summary, our research evidence provides a thorough understanding of the *CmMLO* gene family and demonstrates their potential role in disease resistance, as well as a theoretical foundation for melon disease resistance breeding.

KEYWORDS

***Cucumis melo* L., genome-wide, *MLO* gene family, expression analysis, gene clone**

Introduction

Melon (*Cucumis melo* L.) is an economically important plant with excellent nutritional value (Amanullah et al., 2018; Amanullah et al., 2021). It is a widely distributed fruit crop with a long history of cultivation and has been observed to be susceptible to powdery mildew (PM), especially late in the growing season (Zhang et al., 2013). The pathogens proliferate and spread rapidly in plant parts after the PM infection (Li et al., 2017). In general, white, powdery bacteria cover the leaves, petioles, and stems of infected plants on both sides. In the later stage of infection, the leaves gradually turn yellow, die, and fall off, exposing the fruit to sunlight, which will eventually lead to the overall death of the plant late in growth, seriously affecting the quality and yield of melon (Kuzuya et al., 2003). Mainly, *Podosphaera xanthii (Px)* and *Golovinomyes cichoracearum (Gc)* are the two main fungi that are mainly responsible for the occurrence of PM disease in Cucurbitaceae crops (Křístková et al., 2009).

The Mildew Resistance Locus "O" (MLO) is a gene family unique to plants (Kim et al., 2002). It exists in many plants but is also known as an important member of the *MLO* gene, which plays a negative regulatory role in the process of plant disease resistance, which is equivalent to the "disease susceptibility" gene in plants to a certain extent. This gene was first discovered in barley, and mutants of this gene produce broad-spectrum resistance to barley PM (Schultheiss et al., 2002). Numerous studies have shown that the *MLO* gene has dual functions that are especially involved in a negative regulation of plant disease resistance and a negative regulation of plant leaf cell death (Büschges et al., 1997; Piffanelli et al., 2002). The previously published studies believed that the MLO protein has a complete set of seven transmembrane domains (Büschges et al., 1997; Devoto et al., 1999; Kim et al., 2002). Currently, it has been found that the number of MLO protein transmembrane domains varies in both lower and higher plants. In addition, MLO proteins with <5 transmembrane domains are usually found in lower plants, while MLO proteins with 6–8 transmembrane domains are usually found in higher plants (Chen et al., 2014; Zhang et al., 2018). Some studies have shown that there are 12 to 19 members of the *MLO* gene family in Arabidopsis, grape, rice, peach, woodland strawberry, and tobacco (Devoto et al., 2003; Feechan et al., 2008; Liu and Zhu, 2008; Pessina et al., 2014; Appiano et al., 2015).

In the past few decades, an extensive study of some MLO proteins related to PM resistance has been performed and successfully employed in breeding mechanism of disease resistance. Freisleben (Freisleben and Lein, 1942) treated barley (*Hordeum vulgare*) variety 'Haisa' by X-ray to induce mutagenesis and discovered the first mutant MLO resistant to barley PM. In 1997, further studies revealed that wheat (*Triticum aestivum*) was resistant to PM when the *MLO* gene had mutated into *mlo* (Büschges et al., 1997; Panstruga, 2005b). At present, the *MLO* gene has been implicated in PM sensitivity in *Hordeum vulgare*, *Solanum lycopersicum*, *Arabidopsis thaliana*, *Oryza sativa*, and *T. aestivum* (Elliott et al., 2002). In addition, three mutants, *Hvmlo* (Shirasu et al., 1999), *SImlo* (Bai et al., 2008), and Atmlo2/6/12 triple mutants were found to be resistant to PM (Consonni et al., 2006). In recent years, extensive studies of *MLO* genes revealed that mutations of one or more specific *MLO* genes in *Vitis vinifera*, *Petunia hybrida*, *Pisum sativum*, *Capsicum annuum*, and *T. aestivum* were found to cause broad-spectrum resistance to PM (Humphry et al., 2011; Pavan et al., 2011; Zheng et al., 2013; Wang et al., 2014; Jiang et al., 2016; Pessina et al., 2016), because there is a possibility that PM requires MLO proteins to invade host plant cells (Panstruga and Schulze-Lefert, 2003; Panstruga, 2005b). Therefore, deletion or mutation of the *MLO* gene can cause PM spores to not enter the plant cell wall (Panstruga, 2005b).

In addition, MLO protein plays a negative regulatory role in two independent osmotic PM resistance pathways (Underwood and Somerville, 2008). The synaptic fusion protein PEN1/ROR2 of the first pathway is thought to play a role in regulating vesicle transport (Collins et al., 2003); the direct interaction between MLO and ROR2 was found in yeast hybrid experiments (Panstruga, 2005a); Glucohydrolase PEN2 and ABC transporter PEN3 are involved in another pathway and produce and secrete mycotoxins (Lipka et al., 2005; Stein et al., 2006). Some studies found that the N-terminal of MLO protein was located outside of the cell; moreover, the C-terminal was located inside the cell, and there is a calmodulin-binding domain (CaMBD) of 10–15 amino acids away from the 7th transmembrane domain (Xu and Heath, 1998; Kim et al., 2002; Bhat et al., 2005). It was also proposed that two other conserved domains at the C-terminal of an MLO protein might be related to the sensitivity of a plant to PM (Panstruga and Schulze-Lefert, 2003; Feechan et al., 2008). However, Ca^{2+} also plays an important role in regulation of the plant defense response, and a large number of studies have found that the intracellular Ca^{2+} level increases rapidly when plants are infected by pathogens (Xu and Heath, 1998; Blume et al., 2000). Calmodulin is an important Ca^{2+}-binding protein that plays an important role in the calcium signaling pathway (Yang and Poovaiah, 2003). At present, MLO proteins have been identified in numerous crops (*Hordeum vulgare*, *Oryza sativa*, *Brachypodium distachyon*, *Medicago truncatula*, *Cajanus cajan*, and *Phaseolus vulgaris*), and have been found to contain CaMBD (Liu and Zhu, 2008; Ablazov and Tombuloglu, 2016).

In this study, we carried out a systematic and comprehensive bioinformatics analysis of *CmMLO* gene family members in the melon genome database. Nucleotide sequence information, physicochemical properties, phylogenetic relationships, and expression characteristics of the *CmMLO* gene family were analyzed, which provided the theoretical basis for a further methodical study of the function of each member of the *CmMLO* gene family.

Materials and methods

Plant materials and inoculation treatment

A total of four types of melon lines "PM-resistant lines (MR-1 and PI124112) and PM-susceptible lines (X055 and Topmark)" were selected for experiment materials (Figure 1), and identification of physiological races of PM was done using thirteen general international host identification Cucurbitaceae (Iran H, Topmark, Védrantais, PMR 45, PMR 5, WMR 29, Edisto 47, PI 414723, MR-1, PI 124111, PI124112, PMR 6, and Nantais Oblong). The above materials were provided by the Laboratory of Molecular Genetics and Breeding in Melon, Northeast Agricultural University, Harbin, Heilongjiang, China. All the test materials were grown in nutrient

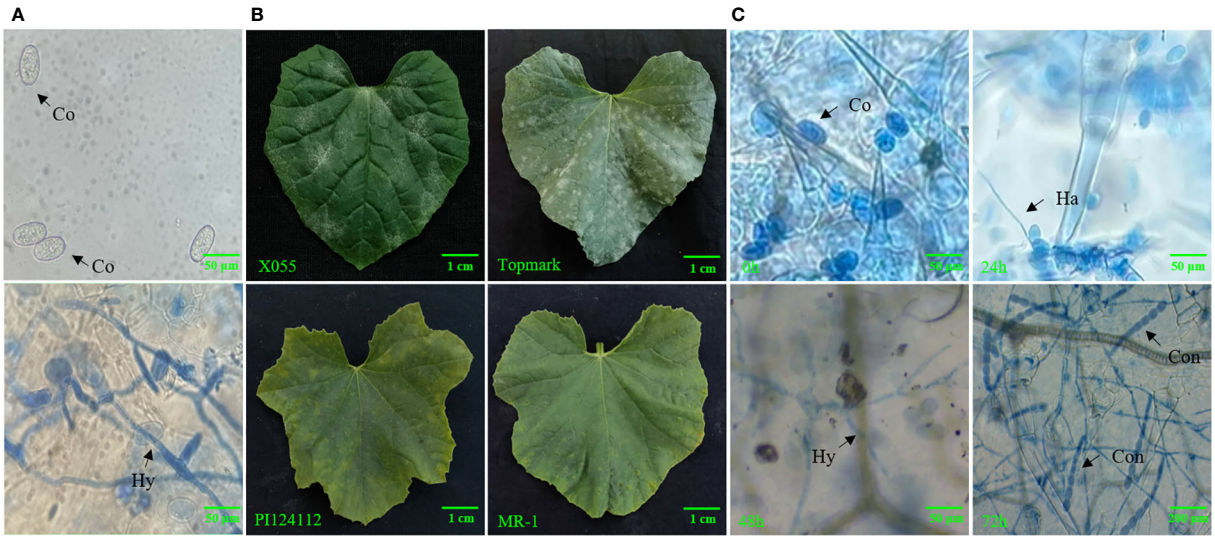


FIGURE 1
(A) Microscopic observation of morphology of PM-susceptible line (Topmark) infected with *Podosphaera xanthii*, **(B)** Phenotypic characteristics of PM-susceptible lines (X055 and Topmark) and PM-resistance lines (PI124112 and MR-1), **(C)** Fungal morphology of PM-susceptible line (Topmark) infected with by *P. xanthii* race 1 at different time interval. Co, conidium; Ha, haustorium; Hy, hyphae; Con, conidiophore.

bowls (9.3 cm in diameter, 9 cm deep) in a controlled environment of greenhouse and kept sterile for several weeks before infection treatment. When the plants reached the three-leaf stage, PM fungal spore suspension spray was used for artificial inoculation, and the concentration of inoculated spores was 1×10^6/mL. Three separate biological replicates were used at each time interval throughout the experiment. Plant leaf tissues were collected at 0, 24, 48, and 72 h after inoculation, snap-frozen in liquid nitrogen, and stored at −80° C prior to the further experiments.

Microscopic observation of PM on melon leaves

The hyphal development was visualized using the Trypan Blue staining technique (Koch and Slusarenko, 1990). In brief, the PM-infected melon leaves were soaked in Trypan Blue staining solution and immediately heated in 100 °C water for 2–5 min. After allowing the solution to cool down to room temperature, the staining solution was discarded and the leaves were decolored using a 2.5 mg/ml chloral hydrate solution, which was replaced every 24 hours until the leaves were completely decolored.

Genome-wide identification of the *MLO* gene family

Firstly, the publicly available genomic data of melon (v3.5.1), watermelon, and cucumber were obtained from the Cucurbitaceae genomic database (http://cucurbitgenomics.org/), the genomic dataset of Arabidopsis was obtained from the The Arabidopsis Information Resource (TAIR) website (https://www.arabidopsis.org/), and bio-informatics analysis were performed as reported earlier (Yang et al., 2022). A total of 15 MLO protein sequences in *A. thaliana* were obtained from the TAIR database, and all the melon proteins were compared with the 15 MLO proteins in *A. thaliana* to obtain the possible gene family members and a bi-directional comparison was performed by using the TBtools software (Chen et al., 2020). Then, all non-redundant protein sequences were analyzed based on further screening, and conservative domains were analyzed to identify the MLO domain using the online website of National Center for Biotechnology Information CD Search (NCBI) (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) and TBtools software. Meanwhile, a total of 14 *MLO* genes obtained from melon were respectively named (*CmMLO1*–*CmMLO14*), according to the earlier reported study (Xu et al., 2014).

Gene structural and conserved motif analysis of the *MLO* genes

The amino acid sequences of the *MLO* gene family in melon were also analyzed by using the online analysis software Multiple Expectation Maximizations for Motif Elicitation (MEME) (http://meme-suite.org/), and ten motifs were identified for subsequent analysis. Meanwhile, a gene structure map was obtained by using the online analysis software of an upgraded gene feature visualization server, "Gene Structure Display Server (GSDS), v2.0" (http://gsds.gao-lab.org/) (Hu et al., 2015).

Phylogenetic tree construction, chromosomal location, and collinearity analysis of the *MLO* genes

The multiple alignments of the amino acid sequences of the *MLO* gene family were analyzed using the ClustalW software (https://www.genome.jp/tools-bin/ClustalW), and defective reading-box sequences and redundant sequences were eliminated manually. The phylogenetic tree of the *MLO* gene family was

constructed using the Neighbor Joining (NJ) method by using the Mega (v7.0) software (http://megasoftware.net) (Chung et al., 2015; Kumar et al., 2016); the bootstrap method with an adjusted value of 1000 was used to evaluate the phylogenetic tree (Kumar et al., 2016), and the constructed tree was enhanced graphically using the itol website (https://itol.embl.de/). The chromosomal location of the *MLO* gene family was extracted from the GFF file of melon data, and a chromosome linkage map was constructed by the TBtools software. Then, the collinearity analysis of *MLO* genes in melon, cucumber, and *A. thaliana* was performed, and Circos maps were generated.

MLO genes cloning in melon

The cloning of the *MLO* genes sequences was performed for PM-resistant lines (MR-1 and PI124112) and PM-susceptible lines (X055 and Topmark), as earlier reported method (Zhang et al., 2023). The information of primer sequences for gene clone can be seen in Supplementary Table S1. The amplification of Polymerase Chain Reaction (PCR) was performed as follows: 1 cycle at 95°C for 5 min, followed by 30 amplification cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 90 s. Then, the amplified targeted fragments were aligned to the vector pMD18-T, and sequencing was performed at Sangon Biotech Co., Ltd.

Expression analysis of *MLO* genes in melon in response to *P. xanthii race*

The infected leaf samples were collected from the four different time intervals of post-inoculation (0 h, 24 h, 48 h, and 72 h). Total RNA was extracted using the Trizol method, and complementary DNA (cDNA) was synthesized through reverse transcription using a ReverTra Ace qPCR RT Kit (Toyobo, Osaka, Japan). The SYBR Green I fluorescence method was used for performing the quantitative real-time polymerase chain reaction (qRT-PCR). The reaction conditions were as follows: pre-denaturation at 94°C for 1 min, denaturation at 94°C for 10 s, annealing at 55°C for 10 s, and extension at 72°C for 20 s. Finally, the obtained qRT-PCR data was processed and analyzed by the $2^{-\Delta\Delta Ct}$ method (Zhang et al., 2021). The information of primer sequences used for qRT-PCR can be seen in Supplementary Table S2.

Statistical data analysis

A total of three biological replicates were performed for each experiment, the numerical datasets were recorded on Microsoft Excel (2016 version), and statistical analysis of descriptive (mean ± standard deviation (x ± s)) statistics was performed using the GraphPad Prism 8.0 software, respectively.

Results

Identification of physiological races of PM

The leaves of PM-infested melons were stained with Trypan blue, and the morphology of the pathogen was observed by light microscopy (Figure 1). It was found that the conidia were around 56–75 µm long and 35–43 µm wide; their shape showed an ovoid shape with a fibrous structure. It was also observed that the mycelium of PM showed linear, curved, and branched types, and that the conidiophores had a long chain structure. By observing the state, length, and width of the conidiophores presented and the presence of fibrils, it was possible to determine that the PM used in this test had the characteristics of *P. xanthii*. Using thirteen international general hosts of Cucurbitaceae as test material and infected with PM, the PM fungus studied in this trial was further identified as *P. xanthii* race 1 based on the resistance (R) and susceptibility (S) of the thirteen internationally identified hosts (Table 1).

Genome-wide identification of *MLO* gene family

In total, 14 candidate genes of the *MLO* family were retrieved from the Cucurbitaceae genome database at different chromosomal locations of whole genome (Table 2). The amino acid (aa) sequences of *MLO* genes were analyzed by Expasy-Protparam, the longest and shortest protein sequences of the *MLO* gene family contained a maximum of 582 and a minimum of 150 amino acids, the molecular weight of *MLO* gene family members in melon ranged from 16.11–66.66 kiloDaltons (kD), the size of the isoelectric point (IP) ranged from 7.25–9.48, and the Introns were ranged from 1–16, respectively. At the same time, TMHMM online software was used to predict the transmembrane domain of the *MLO* gene family, and it was also found that the transmembrane domain of 14 *MLO* genes ranged from 2–7, respectively.

Structural analysis of the *CmMLO* gene family and prediction of conserved motif

The intron distribution provides important evidence related to the phylogenetic relationships among gene family members in plant

TABLE 1 Powdery mildew resistance responses of hosts identification.

Host	Characterization
PMR 45	R
TopMark	S
Vedrantais	S
IranH	S
Nantais Oblong	S
Edisto 47	R
WMR 29	R
MR-1	R
PI 414723	R
PI 124112	R
PI 124111	R
PMR 5	R
PMR 6	R

S is denoting the susceptible and R is representing the resistance characterization, respectively.

10.3389/fpls.2023.1144317

TABLE 2 Identification of fourteen *MLO* genes in melon genome.

Number	Gene ID	Chromosomal location	CDS	Protein length	Introns	TM	PI	Molecular weight/kD
CmMLO1	*MELO3C000169*	Chr0	450	150	1	2	8.71	16.11
CmMLO2	*MELO3C005038*	Chr12	588	195	4	3	9.08	21.51
CmMLO3	*MELO3C005044*	Chr12	1743	580	13	7	9.38	66.66
CmMLO4	*MELO3C007979*	Chr08	1749	582	14	7	9.34	66.37
CmMLO5	*MELO3C012438*	Chr10	1713	570	14	7	9.25	65.35
CmMLO6	*MELO3C013709*	Chr06	1719	572	14	7	8.84	64.58
CmMLO7	*MELO3C013761*	Chr06	1443	480	14	6	7.25	54.85
CmMLO8	*MELO3C016709*	Chr07	1683	560	13	7	9.1	57.23
CmMLO9	*MELO3C019435*	Chr06	1671	556	14	7	9.02	63.73
CmMLO10	*MELO3C021515*	Chr09	1662	553	16	7	8.8	64.01
CmMLO11	*MELO3C022486*	Chr11	1107	368	11	5	9.18	42.22
CmMLO12	*MELO3C023219*	Chr11	1398	465	11	7	9.48	54.87
CmMLO13	*MELO3C025761*	Chr11	1314	437	12	7	9.35	50.25
CmMLO14	*MELO3C026525*	Chr03	1629	542	14	7	8.69	61.59

evolution studies using the online tool "GSDS". The analysis of gene structure showed that the structure of the *MLO* gene family was complex (Figure 2), and introns existed in each gene. The *CmMLO10* gene contained most of the introns, along with 16 introns, while the *CmMLO1* gene had only 1 intron. In terms of gene length, the length of the *CmMLO4* gene was significantly longer when compared to other genes. Seven gene (*CmMLO13, CmMLO2, CmMLO14, CmMLO4, CmMLO6, CmMLO1, CmMLO12*) was found without the UTR region at both ends, and one gene (*CmMLO11*) similarly lacked the UTR region at one end.

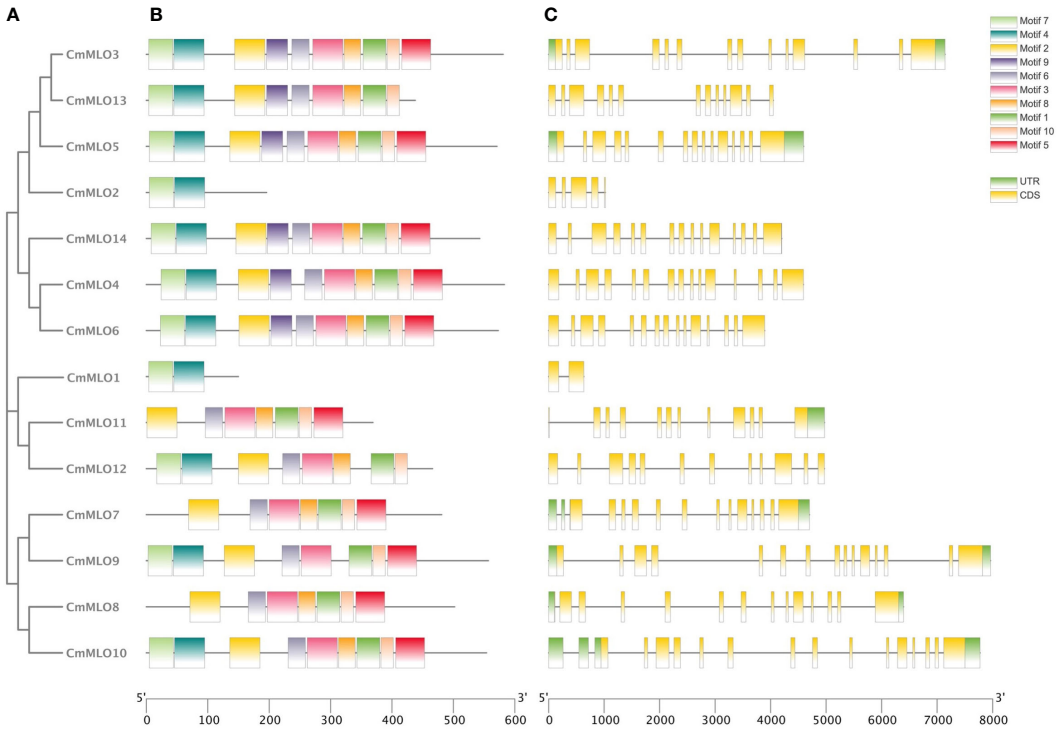


FIGURE 2
Analysis of conserved motif and gene structure of the *CmMLO* gene family. **(A)** phylogenetic tree of the *CmMLO* gene family, **(B)** *CmMLO* gene motif, **(C)** *CmMLO* gene structure.

Gene structural analysis showed that genes in the same clade had similar intron and exon distribution patterns. For example, four genes in the Clade VI group had the same distribution characteristics of the genetic structure.

The amino acid sequences of 14 *CmMLO* genes were analyzed by online MEME software, and ten conserved motifs were obtained. Among them, five genes had ten conserved motifs; *CmMLO11, CmMLO7, CmMLO8* lacked motif 7, and the *CmMLO1* gene and *CmMLO2* had two conserved motifs because the protein sequence of *CmMLO1* and *CmMLO2* was the shortest. The results showed that all of the genes in Clades I did not lack conserved motifs; among them, all four genes in Clade VI had lost some conserved motifs. Meanwhile, we also found that motif 5 was missing in four genes (*CmMLO2, CmMLO1, CmMLO13* and *CmMLO12*) and motif 8 was missing in three genes (*CmMLO2, CmMLO9*, and *CmMLO1*). The results suggested that these conserved motifs are relatively conserved and may play a very important role in the function of the *MLO* gene family. Finally, these findings demonstrated that 14 MLO genes in melon are relatively consistent in conserved motif conservation and have consistent location distribution among some genes. To further understand each conserved motif, we used the MEME online tool to draw a map of motifs 1–10 (Figure 3). The above analysis of the conserved motifs shows that motifs 1 and 4 are highly conserved and these two conserved motifs consist of 38 and 50 amino acid sites, respectively, of which motif 1 has ten completely conserved sites and motif 4 has nine completely conserved sites. The common amino acid conserved site of these two conserved motifs was leucine.

Chromosomal location and collinearity analysis of the *MLO* genes

The positions of these 14 *MLO* genes were identified on their respective chromosomes (Figure 4), among which the *CmMLO1* gene was located on pseudo-chromosome (Chr0) and the other genes were distributed on genetic positions of eight different chromosomes of melon genome; however, three *MLO* genes were found on two chromosomes (Chr6 and Chr11), respectively. A total of four *CmMLO* genes (*CmMLO6*, *CmMLO4*, *CmMLO13*, and *CmMLO2*) were involved in two fragment repeats; among them, *CmMLO6* and *CmMLO4* are fragments of repeating genes from the same clade. To explore the evolutionary relationship of *MLO* genes in different species (*A. thaliana*, *Cucumis melo*, and *C. sativus*), we performed collinearity analyses of genes in three plants (Figure 5). We found 26 pairs of duplicated fragments involving 31 genes in three species. The *CmMLO* and *CsMLO* genes were the most collinear, with 15 pairs in total. This means that these genes have changed little over time and play an important role in PM-resistance in melon and cucumber.

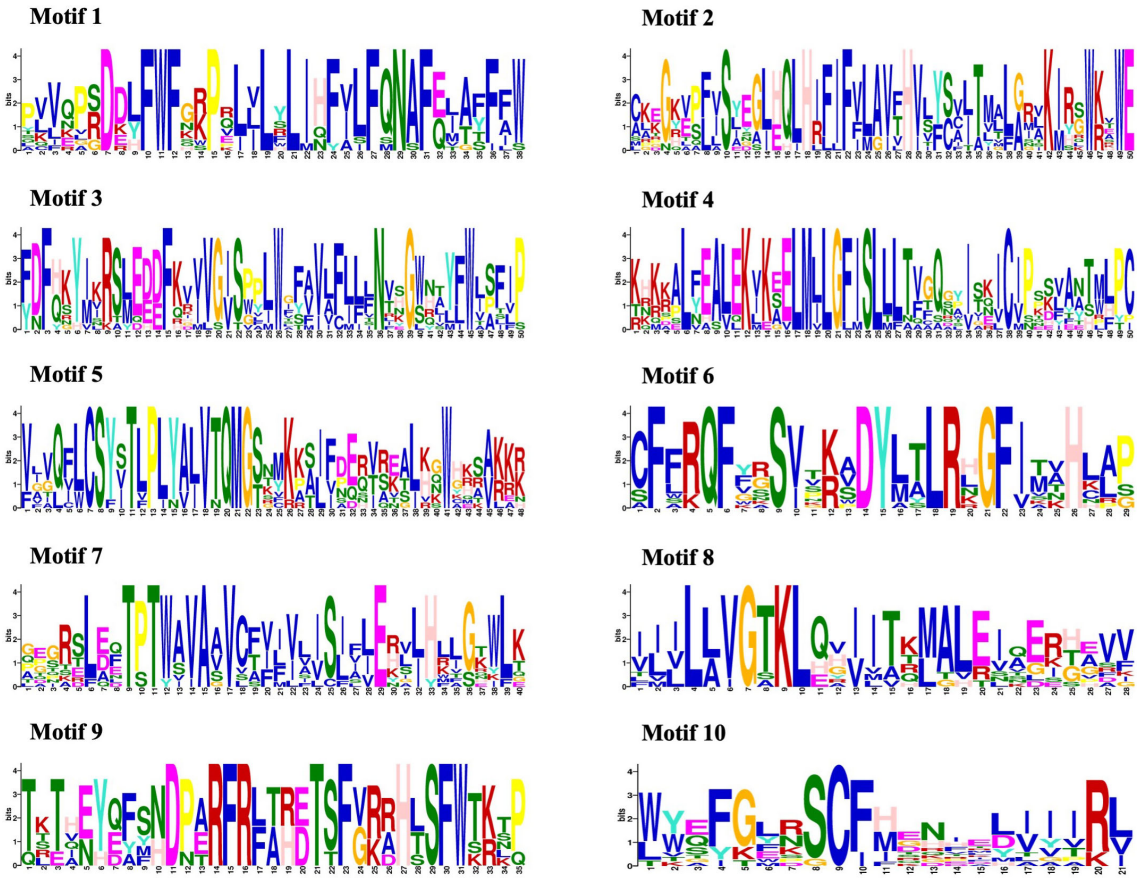


FIGURE 3
Conserved motif analysis of the CmMLO proteins.

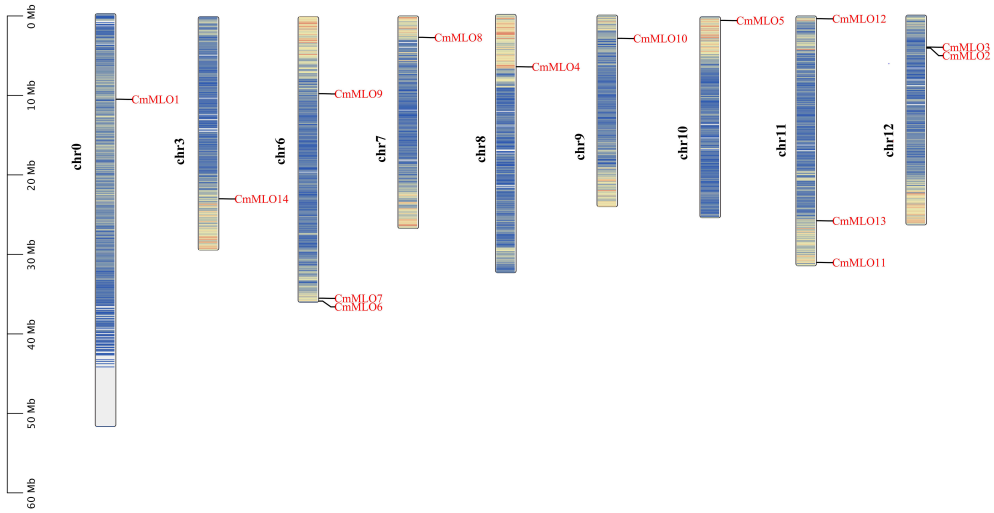


FIGURE 4
Chromosomal distribution of the *CmMLO* gene family members.

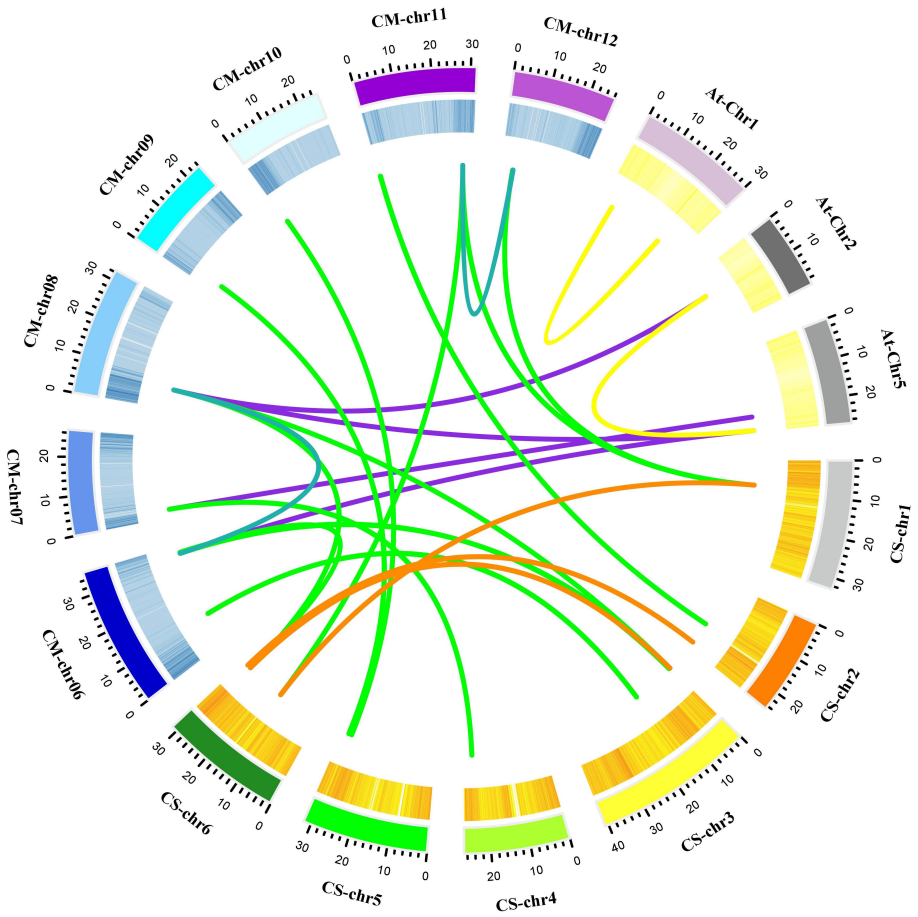


FIGURE 5
Collinearity analysis of the *MLO* gene family in melon. Different line colors indicate the following; green is showing homology between the melon and cucumber genomes; purple is showing homology between the melon and Arabidopsis genomes; blue is showing homology within the melon genome; yellow is showing homology within the Arabidopsis genome; and orange is showing homology within the cucumber genome. These gene pairs are connected by links between chromosomes.

Cloning of the *CmMLO* genes and bioinformatics analysis

To investigate the differences in *MLO* genes in melon lines with different PM resistance, two PM-resistant lines (MR-1 and PI124112) and two PM-susceptible lines (Topmark and X055) were selected as templates for the cloning of the *CmMLO* genes, and a total of thirteen genes were cloned (Supplementary Figure S1). The CDS lengths and amino acid (aa) numbers of the 13 *CmMLO* genes in the four melon lines are shown in Table 3. The transmembrane domain of CmMLO proteins was analyzed by using TMHMM (v2.0), and the results showed that 13 CmMLO proteins have 2–7 TM structures. However, different CmMLO proteins possess different numbers of TM structures, and the transmembrane domains of the same CmMLO protein differ significantly in different melon lines. The CmMLO1, CmMLO2, CmMLO4, CmMLO7, CmMLO10 and CmMLO11 proteins had the same number and position of TM structures in all four melon lines; there were 2, 3, 7, 6, 7, and 5 TM structures, respectively. However, CmMLO3 protein has 3 TM structures in the PM-resistant line MR-1, which is significantly different from the other three melon lines. The TM structures of the CmMLO5 protein in the PM-resistant line (MR-1) significantly differed as compared to the two PM-susceptible lines (X055 and Tompark), with only 5 TM structures in MR-1. However, the position and amount of CmMLO6 protein differed significantly between four melon lines: CmMLO12 protein has 5 TM structures in the PM-resistant line (MR-1) and 7 TM structures in all other melon lines; CmMLO13 protein has 3 TM structures in the PM-resistant line (PI124112), which was significantly different from the other three melon lines.

In order to better study the *CmMLO* gene family, a phylogenetic tree was constructed by combining the CmMLO and MLO proteins of different species (Arabidopsis, wheat, tomato, and pea) (Figure 6). The results showed that the 14 CmMLO proteins obtained from melon genome were divided into different branches, in which CmMLO7, CmMLO9, CmMLO8, and CmMLO10 proteins were distributed with *AtMLO4*, 11 and 14 proteins in Clade VI; CmMLO1, CmMLO12, and CmMLO11 proteins are distributed with AtMLO1, 13, 15 proteins in Clade V; CmMLO6, CmMLO4 and CmMLO14 proteins are distributed with *AtMLO5*, 7, 8, 9, 10 proteins in Clade I; CmMLO13, CmMLO3 and CmMLO5 proteins were distributed with AtMLO2, 6 and 12 proteins, SlMLO1 protein, PsMLO1 protein in Clade III; CmMLO2 protein is distributed with AtMLO3 protein in Clade IV. Although the genes from the same branch are presumed to have similar roles in plant growth and development because of their high similarity of genes from the same branch.

The multiple sequence alignment of the MLO proteins in Clade III (Figure 7) revealed that the CmMLO5 and CmMLO3 proteins in the PM-resistant line (MR-1), the CmMLO13 protein in the PM-resistant line (PI124112) had incomplete and altered transmembrane domain position due to base mutations, while all of the remaining MLO proteins contained 7 TM structures. The analysis of CaMDB and the conserved structural domain revealed that CaMDB had approximately 15-20 amino acids away from the 7th transmembrane domain, conserved structural domains I and II are located at the C-terminus of the MLO protein, and structural domain II has the consistent sequence D/E-F-S/T-F. The CmMLO5 and CmMLO3 proteins in the PM-resistant line MR-1, the CmMLO13 protein in the four melon lines, and the SlMLO1 protein do not have a CaMDB with two conserved domains (I and II), but the remaining MLO proteins all contain a CaMDB and two conserved structural domains (I and II).

Expression characteristics of the *MLO* gene family in melon induced by *P. xanthii race* 1

The changes in *CmMLO* gene expression in four different melon lines infected with *P. xanthii* race 1 were analyzed by real-

TABLE 3 The detailed information of CDS and protein length of the *CmMLO* genes.

Melon lines - genes	CDS	Protein	TM	Melon lines - genes	CDS	Protein	TM
MR-1 - *CmMLO1*	450	150	2	MR-1 - *CmMLO8*	1707	565	6
PI124112 - *CmMLO1*	450	150	2	PI124112 - *CmMLO8*	1707	565	6
X055 - *CmMLO1*	450	150	2	X055 - *CmMLO8*	1707	568	7
Topmark - *CmMLO1*	450	150	2	Topmark - *CmMLO8*	1707	568	7
MR-1 - *CmMLO2*	588	196	3	MR-1 - *CmMLO9*	1610	518	3
PI 124112 - *CmMLO2*	588	196	3	PI 124112 - *CmMLO9*	1610	518	3
X055 - *CmMLO2*	588	196	3	X055 - *CmMLO9*	1671	556	7
Topmark - *CmMLO2*	588	196	3	Topmark - *CmMLO9*	1671	556	7
MR-1 - *CmMLO3*	1741	565	3	MR-1 - *CmMLO10*	1662	552	7
PI 124112 - *CmMLO3*	1743	580	7	PI 124112 - *CmMLO10*	1662	552	7
X055 - *CmMLO3*	1743	580	7	X055 - *CmMLO10*	1662	552	7
Topmark - *CmMLO3*	1743	580	7	Topmark - *CmMLO10*	1662	552	7
MR-1 - *CmMLO4*	1704	567	7	MR-1 - *CmMLO11*	1107	368	5

(Continued)

TABLE 3 Continued

Melon lines - genes	CDS	Protein	TM	Melon lines - genes	CDS	Protein	TM
PI 124112 - *CmMLO4*	1704	567	7	PI 124112 - *CmMLO11*	1107	368	5
X055 - *CmMLO4*	1704	567	7	X055 - *CmMLO11*	1107	368	5
Topmark - *CmMLO4*	1704	567	7	Topmark - *CmMLO11*	1107	368	5
MR-1 - *CmMLO5*	1714	559	5	MR-1 - *CmMLO12*	1397	458	5
PI 124112 - *CmMLO5*	1713	570	7	PI 124112 - *CmMLO12*	1395	464	7
X055 - *CmMLO5*	1713	570	7	X055 - *CmMLO12*	1398	465	7
Topmark - *CmMLO5*	1713	570	7	Topmark - *CmMLO12*	1398	465	7
MR-1 - *CmMLO6*	1718	550	5	MR-1 - *CmMLO13*	1314	437	7
PI124112 - *CmMLO6*	1617	530	7	PI124112 - *CmMLO13*	1313	425	3
X055 - *CmMLO6*	1635	530	5	X055 - *CmMLO13*	1314	437	7
Topmark - *CmMLO6*	1718	557	5	Topmark - *CmMLO13*	1314	437	7
MR-1 - *CmMLO7*	1443	480	6				
PI 124112 - *CmMLO7*	1443	480	6				
X055 - *CmMLO7*	1443	480	6				
Topmark - *CmMLO7*	1443	480	6				

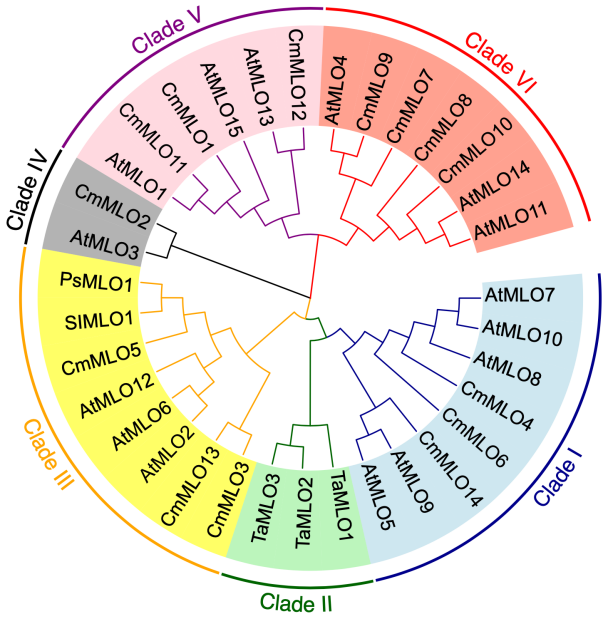


FIGURE 6
Schematic diagram of the *CmMLO* gene family in constructed phylogenetic tree.

time quantitative PCR (Figure 8). We found that the expression changes of 14 *MLO* genes in different resistant and susceptible varieties of melon had great differences after infection with *P. xanthii* race 1. The expression levels of some genes were significantly higher than those of other *MLO* genes in some lines. In addition, significant differences were observed in the expression patterns of different members of the gene family. We found that *CmMLO4* was induced in both PM-susceptible lines at 24 h, while the transcription level was still low at 24 h in resistant materials, and the expression level began to be upregulated at 48 h.

The expression levels of *CmMLO10* and *CmMLO5* were up-regulated at 24 h in a PM-susceptible line, respectively, and no significant change was observed in the two resistant varieties. After infection, nine *CmMLO* genes (*CmMLO7*, *CmMLO8*, *CmMLO12*, *CmMLO6*, *CmMLO13*, *CmMLO3*, and *CmMLO2*) were up-regulated at 24 h in one or two PM-susceptible lines. The

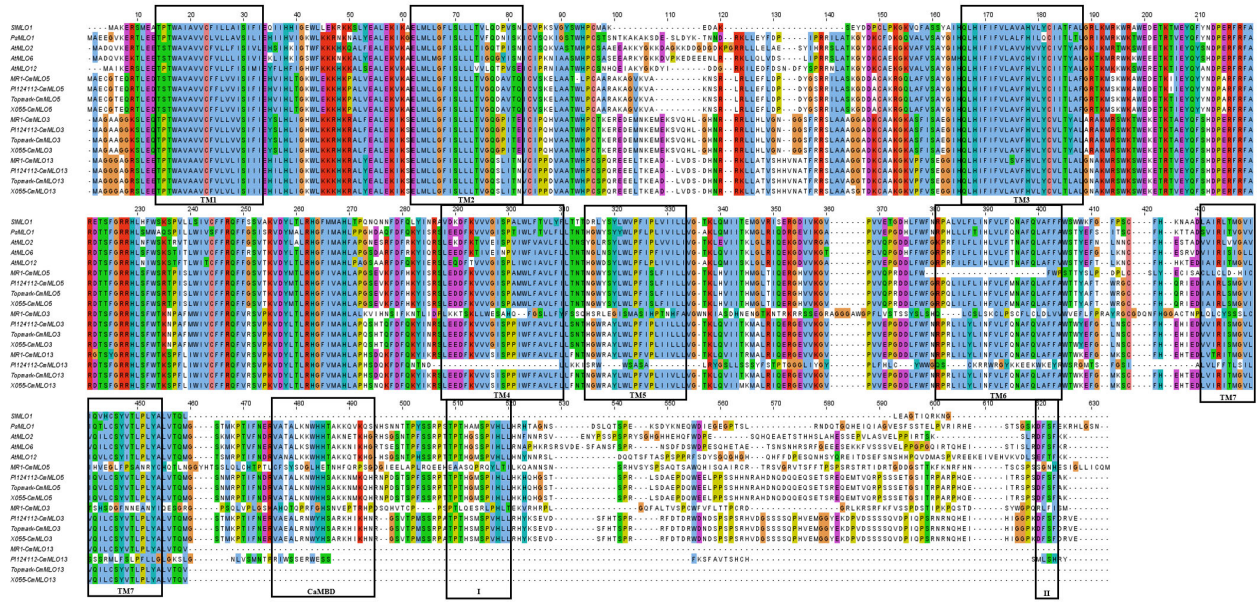


FIGURE 7
Multi-sequence comparison between the CmMLO protein in Clade III and powdery mildew susceptible MLO proteins in PsMLO1, SlMLO1, along with AtMLO2, 6, and 12. TM1–TM7 is representing seven MLO protein transmembrane domains, and I and II representing two conserved domains in the c-terminal of high polymorphism.

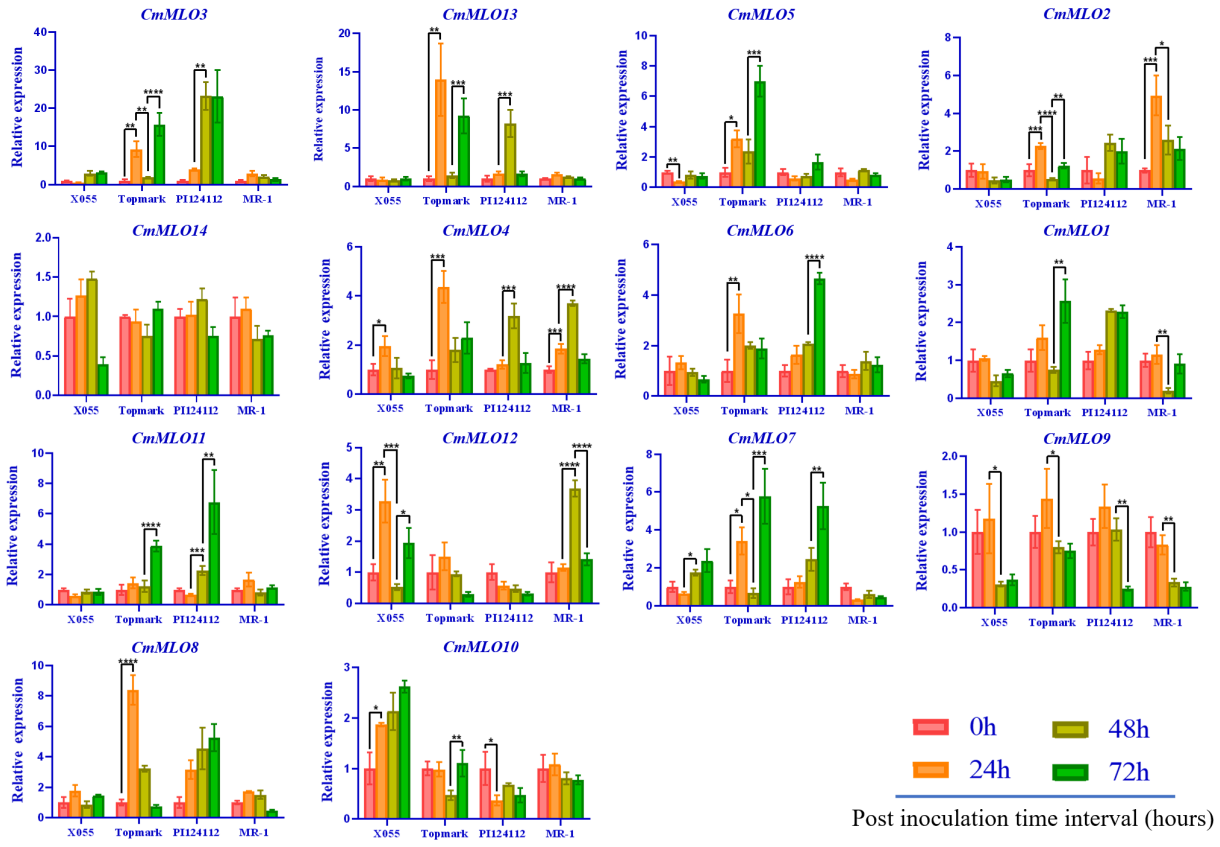


FIGURE 8
The *MLO* genes expression analysis in different melon materials after infection at different time points. *CmActin* was used as an internal control, samples at 0 h after inoculation were used as controls, and the error line is the standard deviation of three biological replicates. Asterick symbols (*, **,***,****) are representing the significant results at **p < 0.01, ***p < 0.001, ****p < 0.001, and *p < 0.05 level, respectively.

expression levels of the other six genes were upregulated at 48 h or 72 h in one of the PM-resistant lines, except for *CmMLO2*; however, the expression trends of these eight genes in the two PM-resistant lines were not completely consistent. In addition, transcription levels of *CmMLO8*, *CmMLO13*, and *CmMLO3* were significantly induced at 24 h in Topmark and 48 h in PI124112. The expression levels of the two lines increased significantly, indicating that the three genes were strongly induced by *P. xanthii* race 1.

After being induced by *P. xanthii* race 1, the expression of *CmMLO3* in the leaves of Topmark and PI124112 was higher than that of the other 13 *CmMLO* genes. At the same time, we also found that the expression level of *CmMLO14* did not change significantly in the four melon lines, and the expression level of *CmMLO9* was down-regulated in the four melon lines after inoculation, suggesting that these two genes may not be induced in the stress response to *P. xanthii* race 1. These results suggest that *CmMLO10* and *CmMLO5* may be genes involved in susceptibility to PM in melon, while *CmMLO8*, *CmMLO4*, *CmMLO13*, and *CmMLO3* may also play a role in PM resistance.

Multiple sequence alignment analysis

We previously hypothesized that three *MLO* genes (*CmMLO13*, *CmMLO3*, and *CmMLO5*) from Clade III might be involved in the regulation of melon PM. We found that the expression of the *CmMLO5* gene was significantly upregulated in the PM-line Topmark after *P. xanthii* race 1 infection. Therefore, the sequences of the *CmMLO5* gene obtained from the sequencing of the four melon lines (MR-1, PI124112, X055, and Topmark) were compared with the reference genome sequence, and 98.4% homology was found, indicating that *CmMLO5* had been successfully cloned.

According to the multiple sequence alignment (Figure 9), there were three base inconsistencies in the nucleotide sequence of the PM-resistant line (MR-1) as compared to the two PM-susceptible lines (X055 and Topmark): a base mutation (G to A) at 279 bp resulting in an amino acid deletion; a base mutation (C to T) at 572 bp resulting in a change of amino acid from T to I; and a base insertion at 1072 bp resulting in a sequence change after amino acid 357th, and sequences of conserved motifs of the *CmMLO* gene family was also altered due to base mutations and insertions. The PM-resistant line PI124112 had one base mutation in the nucleotide sequence as compared to the sequences of the PM-susceptible lines (X055 and Topmark), in common with the PM-resistant line (MR-1), a base mutation (C to T) at 572 bp resulted in a change in amino acid from T to I.

Discussion

In recent years, many studies have shown that the *MLO* gene is significantly associated with PM-resistance in plants, but the in-depth analysis of the *MLO* gene family in melon still remains unclear. In this study, *CmMLO* genes were cloned in four melon lines with different PM-resistances, and the exon-intron composition, lineal homology, and paraphyletic homology of the *MLO* gene family, as well as the changes in expression patterns at different time intervals post-inoculation, were analyzed. A total of 13 *CmMLO* genes, all of which belong to the *MLO* gene family, were cloned from two disease-resistant and two disease-susceptible melon materials. The presence of 14 CmMLO proteins was identified on the basis of the genome-wide gene family analysis, and the absence of the CmMLO14 protein in this study is due to the differences in genetic divergences between the four melon materials and the melon reference genome. The present study results provide a way to sort out how the MLO gene family in melons is related to PM-resistance, and they also give us a theoretical basis for studying its function in more depth.

Further, a total of 14 members of the *CmMLO* gene family were identified by bioinformatic analysis based on melon genome sequencing data. The analysis showed that the conserved motifs and gene structure were different among different members of the *CmMLO* gene family, which further indicated that each member might play different functions. An analysis of the amino acid sites of motifs 1 and 4, two important conserved motifs, showed that leu is the common amino acid site of these two conserved motifs. Leucine may act as a growth regulator or signal molecule used to regulate the biological phenomena and related mechanisms of root growth and morphogenesis (Cheng, 2017). The 3-Methylbdopsyl CoA carboxylase gene is knocked out during Leu degradation, which blocks Leu degradation in mitochondria and affects reproduction in plants (Ding et al., 2012). It has been speculated that *CmMLO* family members containing these conserved motifs may play an important role in promoting growth and development. The chromosomal location analysis showed that the 14 members of the *MLO* gene family were only distributed on eight chromosomes of melon, and the distribution was uneven. There were no *CmMLO* members on four chromosomes: Chr1, Chr2, Chr4, and Chr5. This further indicates the functional diversity of the *CmMLO* gene family.

A phylogenetic tree containing the *MLO* genes from five species was constructed to elucidate the phylogenetic relationships *MLO* family members between *C. melo* and other plants. In total, six different clades (I to VI) were identified (Figure 6), and these clades were named on the basis of classification as earlier reported (Devoto et al., 2003). The phylogenetic analysis divided the *CmMLO* genes into five different clades (Figure 6), which was consistent with the structure, evolution and functional inference on the *MLO* gene family in three cultivated Cucurbitaceae spp as previously reported (Iovieno et al., 2015). The phylogeny analysis showed that the *CmMLO* genes were distributed in five of the six clades. Among them, Clade III includes some *MLO* genes from dicotyledonous plants; these genes were previously determined to be related to PM-susceptibility (Consonni et al., 2006; Bai et al., 2008; Pavan et al., 2011). In Clade III, three *CmMLO* genes (*CmMLO13*, *CmMLO3*, and *CmMLO5*) were clustered with *AtMLO2*, *AtMLO6* and *AtMLO12*, *SlMLO1*, and *PsMLO1*. In addition, the amino acid sequence of the identified *MLO* gene was compared and the conserved sequence of different genes in *Arabidopsis*, tomato, pea, and melon in Clade III were studied. It was found that the three *CmMLO* genes were highly conserved in the predicted TM domain location (Devoto et al., 2003). In addition, CaMBD at the C-terminal of the MLO protein was found to be highly conserved in

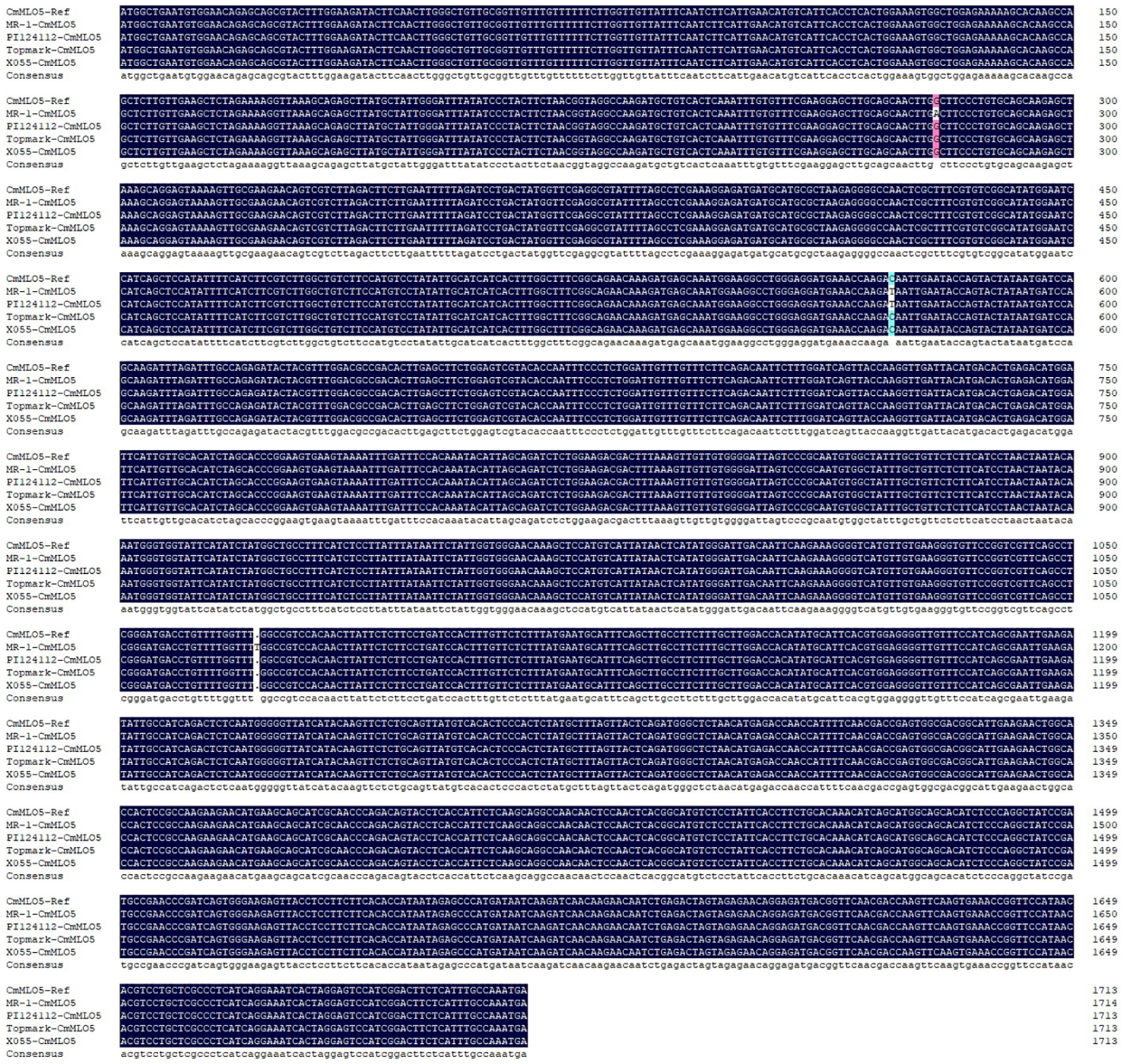


FIGURE 9
Multiple sequence alignment of the *CmMLO5* gene in four melon lines (MR-1, PI124112, X055, and Topmark), respectively.

the entire *Arabidopsis MLO* family (Kim et al., 2002) among the eight different *MLO* gene members analyzed in the above plants. Panstruga and co-researchers also found two other conserved regions (I and II) at the C-terminal of the MLO protein that play an important role in regulating PM infection (Panstruga, 2005b). The peptide domain I is located about 15–20 residues downstream of CaMBD and is characterized by conserved serine and threonine residues. The peptide domain II is located distal to the C-terminal and contains the consistent sequence D/E-F-S/T-F. In this study, the C-terminus of six MLO proteins has these two conserved domains (Figure 7). All of these findings suggest that MLO proteins containing CaMBD and peptide domains I and II are potentially functionally conserved.

According to earlier research findings, *MLO* mutant plants could inhibit the formation of haustrium and epidermal cells infected by the pathogen that did not die immediately (Koga et al., 1990; Nelson et al., 1990; Aist and Bushnell, 1991). In addition, in the broad-spectrum resistance response mediated by the *MLO* gene, the cell wall is thickened by regulating the occurrence of a mastoid reaction to resist the primary infection of PM (Peterhansel et al., 1997). Some researchers have proven through validated experiments that mastoid resistance is controlled by the *MLO* gene. Therefore, the mastoid reaction provides plants with the main defense against pathogen infection of host cells in the early stage of *MLO*-mediated resistance (Aist, 2003). Given that the putative functions of these *MLO* genes in Clade III are derived from *S. lycopersicum*, *P. sativum*, and *A. thaliana*, we infer that this clade is important for *C. melo* because these genes are necessary for PM susceptibility (Consonni et al., 2006; Bai et al., 2008; Pavan et al., 2011). Thus, the clade may be

unique to dicotyledons. The *MLO* gene was previously confirmed to be a gene related to PM sensitivity and the expression level of *this* gene in PM-susceptible plants was rapidly up-regulated after PM infection (Cheng et al., 2013). Herein, we concluded that *CmMLO* genes are associated with melon PM susceptibility and may be significantly up-regulated in susceptible lines following PM infection. The changes in *CmMLO* gene expression in four different melon lines infected with PM were analyzed by real-time quantitative PCR, and we found that the *CmMLO5* gene of Clade III was significantly up-regulated in the susceptible line (Topmark) at 24h and 72h after infection by PM. In combination with the microscopic observation of four periods (0h, 24h, 48h and 72h) of Trypan Blue staining of Topmark leaves of susceptible lines after *P. xanthii* race 1 infection, we hypothesize that the *CmMLO5* gene plays an important role in both periods when the PM pathogen haustorium invades the leaf cells and during hyphal differentiation to produce conidiophores.

Multiple sequence alignment analysis of the *CmMLO5* gene in four melon lines showed that both PM resistant lines (MR-1 and PI124112) showed a single base mutation (C to T) at 572 bp compared to two PM susceptible lines (X055 and Topmark), which resulted in a change in amino acid from T to I. So we speculate that this single base mutation (C to T) at 572 bp, which further results in the loss of protein function and thus confers resistance to PM in melon. However, the expression level of the *CmMLO5* gene in the PM-susceptible line (X055) did not change significantly. This may be caused by the different disease-susceptible mechanisms of different melon lines. In a recent molecular study, a genetic mapping population was developed, and mapping of the recessive gene *CmPMrs* (*MELO3C012438*) was done using a forward genetics mapping strategy, which revealed that stem resistance on melon is mainly controlled by a candidate genetic region positioned on chromosome 10 (Cui et al., 2022). This evidence could also suggest that the *CmMLO5* gene is associated with PM susceptibility.

The specific reasons need to be further tested for verification. The expression level of *CmMLO10* in Clade VI showed up-regulation in susceptible lines after the inoculation, which was consistent with the results of earlier reported study (Dong et al., 2019), but no significant change was observed in the expression level in resistant lines. This may occur because the *MLO* gene is a gene that negatively regulates plant disease resistance (Piffanelli et al., 2002; Kumar et al., 2016), so its expression level in susceptible lines is significantly up-regulated after being infected by *P. xanthii* race 1. However, a great difference was observed in the variation of different PM lines, which inferred that different genes in different varieties had different sensitivities to PM invasion. Mutants of *AtMLO4* and *AtMLO11* in the same clade as *CmMLO10* can regulate asymmetrical root growth (Chen et al., 2006). So, it is thought that *CmMLO10* may also control root growth in addition to making plants resistant to PM.

Further, the expression levels of the *CmMLO8*, *CmMLO13*, and *CmMLO3* genes in resistant and susceptible lines changed significantly, after the induction of *P. xanthii* race 1 infection. In addition, the range of variation in gene expression of *CmMLO3* in the two varieties was larger than that of other genes. After *P. xanthii* race 1 infection, the expression levels of these three genes increased in both PM-resistant and PM-susceptible lines, but the response time was different. However, the up-regulation of *MLO* gene expression in PM-resistant lines may be caused by biological or abiotic stress (Howlader et al., 2017), and the specific reasons need to be further verified by future experiments. After infection with *P. xanthii* race 1, the expression levels of *the CmMLO14* gene at the other three time points in the four lines showed no significant change compared with that at 0 h, indicating that the *CmMLO14* gene was not induced by *P. xanthii* race 1. The expression of the *CmMLO9* gene was down-regulated in the four melon lines at 48 h and 72 h. This may occur because of the inhibition of *CmMLO9* gene expression during the interaction between melon PM and melon, and this type of inhibition does not change due to the natural resistance and susceptibility of melon varieties. So, it is speculated that the *CmMLO9* gene is not involved in the regulation of PM in melon. However, the minimization of the expression level may be caused by other factors or pathogenic bacteria infecting melon indirectly, so it has no direct relationship with pathogenic bacteria. In addition, although two pairs of fragmented genes (*CmMLO6* and *CmMLO4*, and *CmMLO13* and *CmMLO2*) have similarities in their sequences, their expression patterns seem to be different, which indicates that the repeated genes of melon have achieved functional differentiation during the period of evolution.

Conclusion

The current study demonstrated a genome-wide bioinformatics analysis for the identification of the *CmMLO* gene family and found that the amino acid sequence of the *CmMLO* gene shows significant conserved protein characteristics. By cloning the *CmMLO* gene sequence in different melon lines, analyzing the *CmMLO* gene expression pattern after infection, and microscopic observations of the infestation pattern of PM, we concluded that the *CmMLO5* (*MELO3C012438*) gene plays a negative role in regulating PM-resistance in the susceptible line (Topmark). We also assumed that this single base mutation (C to T) at 572 bp further results in the loss of protein function and thus confers resistance to PM in melon. In short, this study is the first step toward figuring out how the *CmMLO* gene family works, which could provide a certain basis for molecular and functional research in an important direction in the future.

Data availability statement

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

Author contributions

TZ and NX performed the bioinformatic analysis and molecular experiment. SA assisted in the data curation, formal analysis, and

scientific writing of the manuscript. SA and PG are the corresponding authors who performed conceptualization, review, and editing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the National Natural Science Foundation of China (NSFC, Grant No. U21A20229 and Grant No. 32250410310 of Research Fund for International Young Scientists (RFIS-I)).

Acknowledgments

We (the authors) are grateful to the Laboratory of Molecular Genetics and Breeding in Watermelon and Melon, for kindly providing the experiment related materials used in this study. All authors contributed to the article and approved the final version.

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.2023.1144317/full#supplementary-material

References

Ablazov, A., and Tombuloglu, H. (2016). Genome-wide identification of the mildew resistance locus O (*MLO*) gene family in novel cereal model species *Brachypodium distachyon*. *Eur. J. Plant Pathol.* 145 (2), 239–253. doi: 10.1007/s10658-015-0833-2

Aist, J. (2003). Papillae and related wound plugs of plant cells. *Annu. Rev. Phytopathol.* 14, 145–163. doi: 10.1146/annurev.py.14.090176.001045

Aist, J. R., and Bushnell, W. R. (1991). "Invasion of plants by powdery mildew fungi, and cellular mechanisms of resistance," in *The fungal spore and disease initiation in plants and animals*. Eds. G. T. Cole and H. C. Hoch (Boston, MA: Springer US), 321–345.

Amanullah, S., Gao, P., Osae, B. A., Saroj, A., Yang, T. , Liu, S., et al. (2021). Genetic linkage mapping and QTLs identification for morphology and fruit quality related traits of melon by SNP based CAPS markers. *Sci. Hortic.* 278, 109849. doi: 10.1016/j.scienta.2020.109849

Amanullah, S., Liu, S., Gao, P., Zhu, Z., Zhu, Q., Fan, C., et al. (2018). QTL mapping for melon (*Cucumis melo* l.) fruit traits by assembling and utilization of novel SNPs based CAPS markers. *Sci. Hortic.* 236, 18–29. doi: 10.1016/j.scienta.2018.02.041

Appiano, M., Pavan, S., Catalano, D., Zheng, Z., Bracuto, V., Lotti, C., et al. (2015). Identification of candidate *MLO* powdery mildew susceptibility genes in cultivated solanaceae and functional characterization of tobacco *NtMLO1*. *Transgenic Res.* 24 (5), 847–858. doi: 10.1007/s11248-015-9878-4

Bai, Y., Pavan, S., Zheng, Z., Zappel, N. F., Reinstädler, A., Lotti, C., et al. (2008). Naturally occurring broad-spectrum powdery mildew resistance in a central American tomato accession is caused by loss of *Mlo* function. *Mol. Plant-Microbe Interact.* 21 (1), 30–39. doi: 10.1094/MPMI-21-1-0030

Bhat, R. A., Miklis, M., Schmelzer, E., Schulze-Lefert, P., and Panstruga, R. (2005). Recruitment and interaction dynamics of plant penetration resistance components in a plasma membrane microdomain. *Proc. Natl. Acad. Sci. U.S.A.* 102 (8), 3135–3140. doi: 10.1073/pnas.0500012102

Blume, B., Nürnberger, T., Nass, N., and Scheel, D. (2000). Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. *Plant Cell* 12 (8), 1425–1440. doi: 10.1105/tpc.12.8.1425

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., et al. (1997). The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* 88 (5), 695–705. doi: 10.1016/s0092-8674(00)81912-1

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

Chen, Z., Hartmann, H., Wu, M., Friedman, E., Chen, J., Pulley, M., et al. (2006). Expression analysis of the *AtMLO* gene family encoding plant-specific seven-transmembrane domain proteins. *Plant Mol. Biol.* 60, 583–597. doi: 10.1007/s11103-005-5082-x

Chen, Y. B., Wang, Y., and Zhang, H. L. (2014). Genome-wide analysis of the mildew resistance locus o (*MLO*) gene family in tomato (*Solanum lycopersicum* l.). *Plant Omics J.* 7 (2), 87–93.

Cheng, X. Y. (2017). *Investigation of biological mechanisms of arabidopsis root system growth and development in response to l-leu and l-lys* (Beijing: China Agric. Univ). doi: 10.27628/d.cnki.gzndu.2017.000071

Cheng, H., Kong, W., Hou, D., Lv, J., and Tao, X. (2013). Isolation, characterization, and expression analysis of *CmMLO2* in muskmelon. *Mol. Biol. Rep.* 40 (3), 2609–2615. doi: 10.1007/s11033-012-2347-8

Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., et al. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for ribo-seq data analysis. *RNA* 21 (10), 1731–1745. doi: 10.1261/rna.052548.115

Collins, N. C., Thordal-Christensen, H., Lipka, V., Bau, S., Kombrink, E., Qiu, J. L., et al. (2003). SNARE-protein-mediated disease resistance at the plant cell wall. *Nature* 425 (6961), 973–977. doi: 10.1038/nature02076

Consonni, C., Humphry, M. E., Hartmann, H. A., Livaja, M., Durner, J., Westphal, L., et al. (2006). Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. *Nat. Genet.* 38 (6), 716–720. doi: 10.1038/ng1806

Cui, H., Fan, C., Ding, Z., Wang, X., Tang, L., Bi, Y., et al. (2022). *CmPMRl* and *CmPMrs* are responsible for resistance to powdery mildew caused by *Podosphaera xanthii* race 1 in melon. *Theor. Appl. Genet.* 135, 1209–1222. doi: 10.1007/s00122-021-04025-4

Devoto, A., Hartmann, H. A., Piffanelli, P., Elliott, C., Simmons, C., Taramino, G., et al. (2003). Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J. Mol. Evol.* 56 (1), 77–88. doi: 10.1007/s00239-002-2382-5

Devoto, A., Piffanelli, P., Nilsson, I., Wallin, E., Panstruga, R., Heijne, G., et al. (1999). Topology, subcellular localization, and sequence diversity of the mlo family in plants. *J. Biol. Chem.* 274, 34993–35004. doi: 10.1074/jbc.274.49.34993

Ding, G., Che, P., Ilarslan, H., Wurtele, E. S., and Nikolau, B. J. (2012). Genetic dissection of methylcrotonyl CoA carboxylase indicates a complex role for mitochondrial leucine catabolism during seed development and germination. *Plant J.* 70 (4), 562–577. doi: 10.1111/j.1365-313X.2011.04893.x

Dong, Y. M., Sun, J. L., Wang, C. Q., Xiao, S. H., Gao, C. , and Jiao, Z. G. (2019). Identification and expression analysis of the syntaxin family genes response to powdery mildew infection in melon. *Acta Hortic. Sin.* 46 (08), 1531–1542. doi: 10.16420/j.issn.0513-353x.2018-0963

Elliott, C., Zhou, F., Spielmeyer, W., Panstruga, R., and Schulze-Lefert, P. (2002). Functional conservation of wheat and rice *Mlo* orthologs in defense modulation to the

powdery mildew fungus. *Mol. Plant Microbe Interact.* 15 (10), 1069–1077. doi: 10.1094/mpmi.2002.15.10.1069

Feechan, A., Jermakow, A. M., Torregrosa, L., Panstruga, R., and Dry, I. B. (2008). Identification of grapevine *MLO* gene candidates involved in susceptibility to powdery mildew. *Funct. Plant Biol.* 35 (12), 1255–1266. doi: 10.1071/fp08173

Freisleben, R., and Lein, A. (1942). Über die auffindung einer mehltauresistenten mutante nach röntgenbestrahlung einer anfälligen reinen linie von sommergerste. *Naturwissenschaften* 30 (40), 608–608. doi: 10.1007/BF01488231

Howlader, J., Park, J. I., Kim, H. T., Ahmed, N. U., Robin, A. H. K., Sumi, K. R., et al. (2017). Differential expression under *Podosphaera xanthii* and abiotic stresses reveals candidate *MLO* family genes in *Cucumis melo* l. *Trop. Plant Biol.* 10 (4), 151–168. doi: 10.1007/s12042-017-9194-7

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

Humphry, M., Reinstädler, A., Ivanov, S., Bisseling, T., and Panstruga, R. (2011). Durable broad-spectrum powdery mildew resistance in pea er1 plants is conferred by natural loss-of-function mutations in *PsMLO1*. *Mol. Plant Pathol.* 12 (9), 866–878. doi: 10.1111/j.1364-3703.2011.00718.x

Iovieno, P., Andolfo, G., Schiavulli, A., Catalano, D., Ricciardi, L., Frusciante, L., et al. (2015). Structure, evolution and functional inference on the *Mildew locus o* (*MLO*) gene family in three cultivated cucurbitaceae spp. *BMC Genom.* 16, 1112. doi: 10.1186/s12864-015-2325-3

Jiang, P., Chen, Y., and Wilde, H. D. (2016). Reduction of *MLO1* expression in petunia increases resistance to powdery mildew. *Sci. Hortic.* 201, 225–229. doi: 10.1016/j.scienta.2016.02.007

Kim, M. C., Lee, S. H., Kim, J. K., Chun, H. J., Choi, M. S., Chung, W. S., et al. (2002). Mlo, a modulator of plant defense and cell death, is a novel calmodulin-binding protein. isolation and characterization of a rice mlo homologue. *J. Biol. Chem.* 277 (22), 19304–19314. doi: 10.1074/jbc.M108478200

Koch, E., and Slusarenko, A. (1990). *Arabidopsis* is susceptible to infection by a downy mildew fungus. *Plant Cell* 2 (5), 437–445. doi: 10.1105/tpc.2.5.437

Koga, H., Bushnell, W., and Zeyen, R. (1990). Specificity of cell type and timing of events associated with papilla formation and the hypersensitive reaction in leaves of *Hordeum vulgare* attacked by *Erysiphe graminis* f.sp. *hordei*. *Can. J. Plant Sci.* 68, 2344–2352. doi: 10.1139/b90-299

Křístková, E., Lebeda, A., and Sedláková, B. (2009). Species spectra, distribution and host range of cucurbit powdery mildews in the Czech republic, and in some other European and middle Eastern countries. *Phytoparasitica* 37 (4), 337–350. doi: 10.1007/s12600-009-0045-4

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

Kuzuya, M., Hosoya, K., Yashiro, K., Tomita, K., and Ezura, H. (2003). Powdery mildew (*Sphaerotheca fuliginea*) resistance in melon is selectable at the haploid level. *J. Exp. Bot.* 54 (384), 1069–1074. doi: 10.1093/jxb/erg100

Li, B., Zhao, Y., Zhu, Q., Zhang, Z., Fan, C., Amanullah, S., et al. (2017). Mapping of powdery mildew resistance genes in melon (*Cucumis melo* l.) by bulked segregant analysis. *Sci. Hortic.* 220, 160–167. doi: 10.1016/j.scienta.2017.04.001

Lipka, V., Dittgen, J., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., et al. (2005). Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science* 310 (5751), 1180–1183. doi: 10.1126/science.1119409

Liu, Q., and Zhu, H. (2008). Molecular evolution of the *MLO* gene family in *Oryza sativa* and their functional divergence. *Gene* 409 (1-2), 1–10. doi: 10.1016/j.gene.2007.10.031

Nelson, H., Shiraishi, T., and Oku, H. (1990). Primary infection of barley by *Erysiphe graminis* f. sp. hordei in relation to leaf-age dependent resistance and the roles of the epidermis and mesophyll in this resistance. *J. Phytopathol.* 128, 55–61. doi: 10.1111/j.1439-0434.1990.tb04251.x

Panstruga, R. (2005a). Discovery of novel conserved peptide domains by ortholog comparison within plant multi-protein families. *Plant Mol. Biol.* 59 (3), 485–500. doi: 10.1007/s11103-005-0353-0

Panstruga, R. (2005b). Serpentine plant MLO proteins as entry portals for powdery mildew fungi. *Biochem. Soc Trans.* 33 (2), 389–392. doi: 10.1042/bst0330389

Panstruga, R., and Schulze-Lefert, P. (2003). Corruption of host seven-transmembrane proteins by pathogenic microbes: a common theme in animals and plants? *Microbes Infect.* 5 (5), 429–437. doi: 10.1016/S1286-4579(03)00053-4

Pavan, S., Schiavulli, A., Appiano, M., Marcotrigiano, A. R., Cillo, F., Visser, R. G. F., et al. (2011). Pea powdery mildew er1 resistance is associated to loss-of-function mutations at a *MLO* homologous locus. *Theor. Appl. Genet.* 123 (8), 1425–1431. doi: 10.1007/s00122-011-1677-6

Pessina, S., Lenzi, L., Perazzolli, M., Campa, M., Dalla Costa, L., Urso, S., et al. (2016). Knockdown of *MLO* genes reduces susceptibility to powdery mildew in grapevine. *Hortic. Res.* 3 (1), 16016. doi: 10.1038/hortres.2016.16

Pessina, S., Pavan, S., Catalano, D., Gallotta, A., Visser, R. G., Bai, Y., et al. (2014). Characterization of the *MLO* gene family in rosaceae and gene expression analysis in *Malus domestica*. *BMC Genom.* 15 (1), 618. doi: 10.1186/1471-2164-15-618

Peterhansel, C., Freialdenhoven, A., Kurth, J., Kolsch, R., and Schulze-Lefert, P. (1997). Interaction analyses of genes required for resistance responses to powdery mildew in barley reveal distinct pathways leading to leaf cell death. *Plant Cell.* 9 (8), 1397–1409. doi: 10.1105/tpc.9.8.1397

Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., et al. (2002). The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress Stimuli1. *Plant Physiol.* 129 (3), 1076–1085. doi: 10.1104/pp.010954

Schultheiss, H., Dechert, C., Kogel, K.-H., and Hückelhoven, R. (2002). A small GTP-binding host protein is required for entry of powdery mildew fungus into epidermal cells of barley. *Plant Physiol.* 128 (4), 1447–1454. doi: 10.1104/pp.010805

Shirasu, K., Nielsen, K., Piffanelli, P., Oliver, R., and Schulze-Lefert, P. (1999). Cell-autonomous complementation of mlo resistance using a biolistic transient expression system. *Plant J.* 17 (3), 293–299. doi: 10.1046/j.1365-313X.1999.00376.x

Stein, M., Dittgen, J., Sánchez-Rodríguez, C., Hou, B., Molina, A., Schulze-Lefert, P., et al. (2006). Arabidopsis PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell* 18 (3), 731–746. doi: 10.1105/tpc.105.038372

Underwood, W., and Somerville, S. C. (2008). Focal accumulation of defences at sites of fungal pathogen attack. *J. Exp. Bot.* 59 (13), 3501–3508. doi: 10.1093/jxb/ern205

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32 (9), 947–951. doi: 10.1038/nbt.2969

Xu, J., Chen, X. Z., Wang, Y., Shi, J. L., Zhu, L. J., and Wang, K. L. (2014). Comparative genomic analysis of MLO gene families in cucumber, melon and watermelon. *J. Nuc. Agric. Sci.* 28 (06), 1006–1017. doi: 10.11869/j.issn.100-8551.2014.06.1006

Xu, H., and Heath, M. C. (1998). Role of calcium in signal transduction during the hypersensitive response caused by basidiospore-derived infection of the cowpea rust fungus. *Plant Cell* 10 (4), 585–597. doi: 10.1105/tpc.10.4.585

Yang, T., and Poovaiah, B. W. (2003). Calcium/calmodulin-mediated signal network in plants. *Trends Plant Sci.* 8 (10), 505–512. doi: 10.1016/j.tplants.2003.09.004

Yang, T., Zhang, P., Pan, J., Amanullah, S., Luan, F., Han, W., et al. (2022). Genome-wide analysis of the peroxidase gene family and verification of lignin synthesis-related genes in watermelon. *Int. J. Mol. Sci.* 23, 642. doi: 10.3390/ijms23020642

Zhang, T., Cui, H., Luan, F., Liu, H., Ding, Z., Amanullah, S., et al. (2023). A recessive gene *Cmpmr2F* confers powdery mildew resistance in melon (*Cucumis melo* l.). *Theor. Appl. Genet.* 136 (1), 1–16. doi: 10.1007/s00122-023-04269-2

Zhang, T., Liu, J., Amanullah, S., Ding, Z., Cui, H., Luan, F., et al. (2021). Fine mapping of *Cla015407* controlling plant height in watermelon. *J. Am. Soc Hortic.* 146 (3), 196–205. doi: 10.21273/JASHS04934-20

Zhang, C., Ren, Y., Guo, S., Zhang, H., Gong, G., Du, Y., et al. (2013). Application of comparative genomics in developing markers tightly linked to the *Pm-2F* gene for powdery mildew resistance in melon (*Cucumis melo* l.). *Euphytica* 190 (2), 157–168. doi: 10.1007/s10681-012-0828-4

Zhang, X. L., Zhang, J. S., Lei, B., Yu, J., and Zhao, D. G. (2018). Research progress of *MLO* proteins in plants. *Plant Physiol. J.* 54, 1159–1171. doi: 10.13592/j.cnki.ppj.2018.0160

Zheng, Z., Nonomura, T., Appiano, M., Pavan, S., Matsuda, Y., Toyoda, H., et al. (2013). Loss of function in *Mlo* orthologs reduces susceptibility of pepper and tomato to powdery mildew disease caused by *Leveillula taurica*. *PloS One* 8 (7), e70723. doi: 10.1371/journal.pone.0070723

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 March 2023
DOI 10.3389/fpls.2023.1142856

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Yong Xu,
Beijing Academy of Agriculture and Forestry Sciences, China
Md Abdur Rahim,
Sher-e-Bangla Agricultural University, Bangladesh

*CORRESPONDENCE
Ah-Young Shin
shinay@kribb.re.kr
Yi Lee
leeyi22@hanmail.net
Suk-Yoon Kwon
sykwon@kribb.re.kr

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

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

RECEIVED 12 January 2023
ACCEPTED 15 February 2023
PUBLISHED 01 March 2023

CITATION
Nie H, Kim M, Lee S, Lim S, Lee MS, Kim JH, Noh SJ, Park SW, Kim S-T, Shin A-Y, Lee Y and Kwon S-Y (2023) High-quality genome assembly and genetic mapping reveal a gene regulating flesh color in watermelon (*Citrullus lanatus*). *Front. Plant Sci.* 14:1142856. doi: 10.3389/fpls.2023.1142856

High-quality genome assembly and genetic mapping reveal a gene regulating flesh color in watermelon (*Citrullus lanatus*)

Hualin Nie[1†], Moonkyo Kim[2,3†], Sanghee Lee[1,4], Sohee Lim[1], Mi Sun Lee[2], Ju Hyeok Kim[2], Sol Ji Noh[5], Seong Won Park[5], Sang-Tae Kim[6], Ah-Young Shin[1,7*], Yi Lee[2*] and Suk-Yoon Kwon[1,4*]

[1]Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea, [2]Department of Industrial Plant Science and Technology, Chungbuk National University, Cheongju, Republic of Korea, [3]Division of Life Science, Korea Polar Research Institute, Incheon, Republic of Korea, [4]Biosystems and Bioengineering Program, Korea Research Institute of Bioscience and Biotechnology (KRIBB) School of Biotechnology, University of Science and Technology, Daejeon, Republic of Korea, [5]Watermelon and Strawberry Research Institute, Chungcheongbuk-do Agricultural Research and Extension Services, Cheongju, Republic of Korea, [6]Department of Medical and Biological Sciences, Catholic University of Korea, Bucheon, Republic of Korea, [7]Department of Bioinformatics, Korea Research Institute of Bioscience and Biotechnology (KRIBB) School of Bioscience, University of Science and Technology, Daejeon, Republic of Korea

The unique color and type characteristics of watermelon fruits are regulated by many molecular mechanisms. However, it still needs to be combined with more abundant genetic data to fine-tune the positioning. We assembled genomes of two Korean inbred watermelon lines (cv. 242-1 and 159-1) with unique color and fruit-type characteristics and identified 23,921 and 24,451 protein-coding genes in the two genomes, respectively. To obtain more precise results for further study, we resequenced one individual of each parental line and an F_2 population composed of 87 individuals. This identified 1,539 single-nucleotide polymorphisms (SNPs) and 80 InDel markers that provided a high-density genetic linkage map with a total length of 3,036.9 cM. Quantitative trait locus mapping identified 15 QTLs for watermelon fruit quality-related traits, including β-carotene and lycopene content in fruit flesh, fruit shape index, skin thickness, flesh color, and rind color. By investigating the mapping intervals, we identified 33 candidate genes containing variants in the coding sequence. Among them, Cla97C01G008760 was annotated as a phytoene synthase with a single-nucleotide variant (A → G) in the first exon at 9,539,129 bp of chromosome 1 that resulted in the conversion of a lysine to glutamic acid, indicating that this gene might regulate flesh color changes at the protein level. These findings not only prove the importance of a phytoene synthase gene in pigmentation but also explain an important reason for the color change of watermelon flesh.

KEYWORDS

watermelon, genome assembly, linkage map, quantitative trait locus (QTL), carotenoid biosynthesis, phytoene synthase

Introduction

Watermelon (*Citrullus lanatus*, 2n = 2x = 22), originally from South Africa, is an edible fruit crop belonging to the family Cucurbitaceae (Bae et al., 2020) that is rich in lycopene, citrulline, vitamin C, and other essential micronutrients and vitamins (Dubey et al., 2021). Humans have domesticated and grown watermelons for 4,000 years (Paris, 2015). Although the original wild watermelon had small fruit (about 0.2 kg) with watery, hard-textured, pale-colored, and bland or bitter-tasting flesh, common watermelons have become sweet and savory after centuries of improvement (Paris, 2015; Guo et al., 2019; Chomicki et al., 2020). There are now more than 1,200 watermelon varieties worldwide with fruits of various sizes, shapes, rind patterns, and flesh colors (Perkins-Veazie et al, 2012; Porcher et al., 2013). According to the characteristics of their fruit, watermelons can be divided into three cultivar groups: *Citroides* ("Red-Seeded" preserving melon, red-seeded citron), *Lanatus* (Tsamma, Kalahari, South-African, and wild watermelons), and *Vulgaris* (commonly cultivated watermelon) (Guo et al., 2019).

The color of watermelon flesh is closely related to the carotenoid composition and can be defined as white, pale yellow, canary yellow, salmon yellow, orange, crimson red (red), scarlet red, or green (Zhao et al., 2013). In watermelons with red, pink, or scarlet flesh, lycopene accounts for 70%–90% of the total carotenoids (Perkins-Veazie et al., 2006; Nadeem et al., 2022). Watermelons with yellow flesh have high levels of xanthophylls (mainly neoxanthin, violaxanthin, and neochrome) (Bang et al., 2010; Liu et al., 2012). Watermelons with orange flesh have much more β-carotene, ζ-carotene, and prolycopene than other pigments (Tadmor et al., 2005). White-flesh watermelons have almost no pigment and only contain trace amounts of phytofluene (Wang et al., 2019). These differences in carotenoid composition are likely due to molecular variation in the processes of carotenoid biosynthesis and accumulation. In addition to flesh colors, a variety of watermelon rind colors are highly desirable to consumers, and breeders place considerable emphasis on this trait. The rind coloration of watermelon is affected by chlorophylls and carotenoids that are synthesized in plant plastids from metabolic precursors provided by the methylerythritol 4-phosphate (MEP) pathway (Simpson et al., 2016; Liu et al., 2020; Ma et al., 2021). The rind colors are likely to be influenced by the same or similar molecular mechanisms that influence the flesh colors (Liu et al., 2020; Xiao et al., 2022). Other traits that are important for watermelon quality include the shape, weight, firmness, and diameter of the fruit. Many previous studies have applied genetic mapping and gene prediction to investigate these unique traits in watermelons (Sandlin et al., 2012; Guo et al., 2015; Liu et al., 2020; Sun et al., 2020).

Several watermelon genomes have been assembled for functional genomics studies and genetic improvement. A non-bitter, white-fleshed Sudanese Kordofan melon (*C. lanatus* subsp. *Cordophanus*) assembled in 2019 appears to be the most closely related ancestor of the domesticated watermelon (Renner et al., 2021). The first high-quality assembled genome sequence of domesticated watermelon was released in 2013 as a result of comprehensive genomic and transcriptomic analyses of the East Asian watermelon cultivar 97103, which contains preferentially selected genomic regions and has lost many disease-resistant genes during domestication (Guo et al., 2013). Subsequently, Wu et al. (2019) created a high-quality genome assembly of a principal American cultivar, Charleston Gray, and conducted a comparative analysis of the Charleston Gray and 97103 genomes. Deng et al. (2022) recently completed assembly of the telomere-to-telomere gap-free genome G42 using high-coverage and accurate long-read sequencing data. These genomic resources were developed to study the evolution, domestication, and genotyping of watermelons. In addition, resequencing analyses have further improved capabilities to identify single-nucleotide polymorphisms (SNPs), simple sequence repeats, and InDels that can be used as co-dominant markers to determine the genotypes of watermelon populations (Rhee et al., 2015).

Watermelon cultivars 242-1 and 159-1 have been bred in the Chungcheongbuk-do Agricultural Research and Extension Services (Cheongju, Korea), which have unique characteristics of color and fruit type: 242-1 has orange flesh, black rind, and oval shape, whereas 159-1 has red flesh, yellow rind, and round shape. To provide more genetic information about watermelon cultivars with unique flesh colors, we assembled the genomes of cultivars 242-1 and 159-1 using Nanopore long-read sequencing and Illumina short-read sequencing. We then resequenced the whole genomes of F_2 **population** to construct a high-density genetic map and perform quantitative trait locus (QTL) analysis. Finally, we investigated the mapped QTL regions to identify candidate genes associated with flesh quality-related traits, including lycopene, β-carotene, flesh color, skin color, skin thickness, and fruit shape index. Our results provide a basis for candidate gene selection, annotation, and functional verification to elucidate the genetics of valuable traits and contribute to the development of breeding strategies based on molecular markers.

Materials and methods

Plant materials and DNA sequencing

We selected the Korean inbred watermelon cultivars 242-1 and 159-1 as experimental materials because of their differential characteristics (Figure 1). Seeds of the 242-1 and 159-1 cultivars were kindly provided by Chungcheongbuk-do Agricultural Research and Extension Services (ASES; Cheongju, Korea). For DNA sequencing, seeds were planted in the ASES experimental field and harvested on day 80 of cultivation.

For genome assembly, high–molecular weight genomic DNA was isolated from the fresh young leaf tissues of 242-1 and 159-1 plants by performing a nuclei isolation step according to previous protocols (Zerpa-Catanho et al., 2021). Libraries for long-read sequencing were prepared by end-repair and dA-tailing, ligation, and purified-ligated DNA purification using NEBNext® Ultra™ II End Repair/dA-Tailing Module (cat. no. E7546, New England Biolabs, Ipswich, MA, USA), NEBNext® Quick Ligation Module (cat. no. E6056, New England Biolabs), and Ligation Sequencing Kit

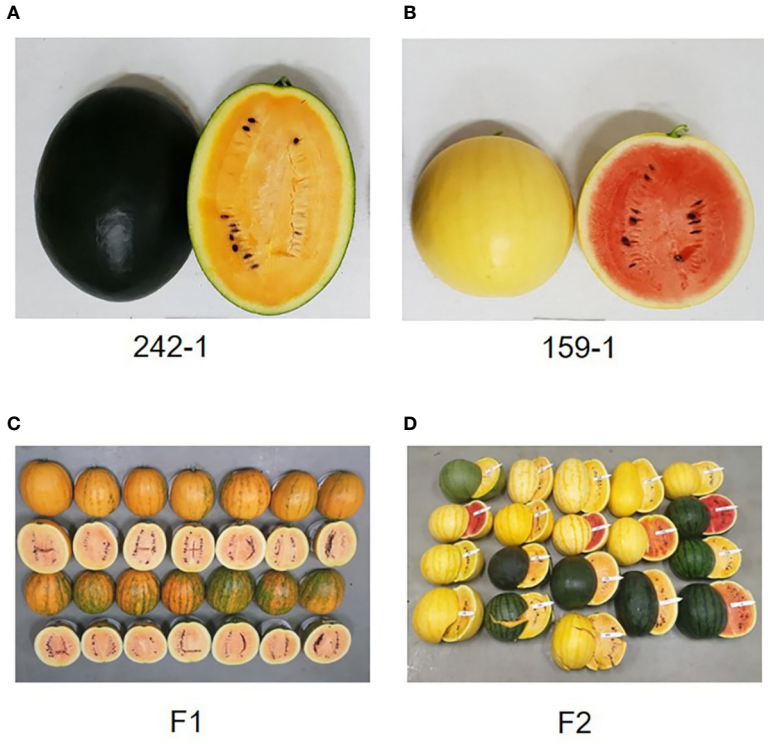


FIGURE 1
Morphological characteristics of the inbred parental lines and F_1 and F_2 individuals. **(A, B)** Fruits of parental cultivars 242-1 **(A)** and 159-1 **(B)**. **(C)** Fruits of F_1 hybrids are round with striped rind and yellow flesh color. **(D)** Fruits of F_2 individuals produced by self-pollination of F_1 individuals. The F_2 generation segregated into various rind colors (dark green, black, and mixture of black and dark green) and flesh colors (red, orange, and mixture of red and orange).

[cat. no. SQK-LSK109, Oxford Nanopore Technologies Co. (ONT), Oxfordshire, UK], respectively, according to recommendations by ONT. MinION sequencing was performed as per the manufacturer's guidelines using R9.4 SpotON Flow Cell (cat. no. FLO-MIN106) and controlled using ONT MinKOW software (ONT). For short-read sequencing, genomic DNA paired-end (PE) and mate-pair (MP) libraries were constructed with 350–800, 550–800, and 600–800 bp insert sizes using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs) and with 2- and 5-kb insert sizes using the Nextera Mate Pair Kit (Illumina, San Diego, CA, USA), respectively. The PE and MP libraries were sequenced at DNACARE Co. (Seoul, Korea)

Genome assembly and pseudomolecule construction

We used the Trimmomatic v0.38 tool to check the quality of the Nanopore long-read sequences and delete the adapters, low-quality reads (reads that contained "N" as more than 10% of the nucleotides), and duplicated reads. The clean PE reads were assembled using the NextDenovo v2.5.0 software with the default parameters. The reads were mapped to an initial assembly using Minimap2 v2.17 with the options "–MD -ax map-ont -t 20 -L" and then sorted using SAMtools v1.16.1 software (Li et al., 2009). The raw assembly was polished with the short-reads from Illumina sequencing in three rounds using NextPolish v1.4.0 software (Hu et al., 2020). Then, haplotigs were removed using PurgeHaplotigs (Roach et al., 2018) with the default parameters. Finally, we validated the genome assemblies using the Benchmarking Universal Single-Copy Ortholog (BUSCO) v4.1.4 software (Simão et al., 2015) with 1,614 Nb of BUSCO markers in Embryophyta (odb10).

The assembled genome sequences were separated into chromosomal units corresponding to the reference sequence for *C. lanatus* cultivar 97103 (http://cucurbitgenomics.org/organism/1) using the NUCMER 4 program (Marçais et al., 2018). Anchoring was performed to determine the order and orientation of sequences within the chromosome. Pseudomolecules were constructed using an in-house perl script, and 100 N's were artificially inserted between contigs to complete the sequence of each chromosome unit. The assembled pseudomolecules were then used for nucmer alignment with the 97103 reference genome, and a dot plot was generated using mummerplot to finally confirm the completeness of the pseudomolecules.

Genome annotation, gene prediction, and functional annotation

The process of genome annotation was divided into three parts: repeat annotation, structural annotation, and functional annotation

(Supplementary Figure S1). Repeat annotation was performed using a homology-based method and a *de novo* method in parallel. Homology-based repeat annotation was performed using RepeatMasker v4.0.3 (Tarailo-Graovac and Chen, 2009) with the plant repeat database PlantsRep (http://www.plantrep.cn/) (Luo et al., 2022). For *de novo* repeat annotation, RepeatModeler v1.0.8 (Flynn et al., 2020) was used to predict *de novo* repeats from the assembled genome sequence and construct a RepeatMasker library for further analysis. The results of the two repeat analyses were integrated, and the repetitive regions in the genome sequences were masked using bedtools maskfasta (Quinlan and Hall, 2010).

We downloaded and analyzed the Illumina short-read transcripts of five tissues (flesh, SRR10389406; leaf, SRR3156561; light green rind, SRR10803496; root, SRR12079410; and shoot apex, SRR3156569) of watermelon (*Citrullus lanantus*) from NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) (Wang et al., 2021). We trimmed the raw RNA sequence using the trimmomatic v0.39 to obtain transcript sequences with an average phred quality score of at least 30 and a minimum length of 50 bp. By removing the Illumina adapter sequences, a total of 27 Gb of paired high-quality nucleotide sequences were obtained for annotation. Gene model predictions for structural annotation were made using the PASA v2.5.1 program (Haas et al, 2003) based on *de novo* transcript assemblies generated with the Trinity v2.8.6 program (Haas et al, 2013) and mapped-region information for RNA-seq short-read data generated with the StringTie v2.0.4 program (Kovaka et al., 2019). In addition, 22,596 protein sequences from the 97103 reference genome were mapped by the protein2genome method in the Exonerate v2.2.0 program and used for gene prediction (Slater and Birney, 2005). Finally, all prediction results (*ab initio*, transcript-based, and protein alignment) were merged to generate a non-redundant original gene set in Evidence Modeler (Haas et al., 2008).

Functional annotation of the gene set was performed by querying the protein sequences against the RefSeq (4,416,162 sequences; https://www.ncbi.nlm.nih.gov/refseq/), UniProt (557,491 sequences; https://www.uniprot.org), and The Arabidopsis Information Resource (TAIR) (48,359 sequences; https://www.arabidopsis.org) databases with an e-value cutoff of 1e−4. Analyses of conserved protein domains, gene ontology, and pathways were performed using the InterProScan program (Jones et al., 2014) and the Pfam, Gene Ontology, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Gene expression information was obtained for regions with FPKM values greater than 0.5 in the integrated gene model based on the watermelon RNA-seq maps.

Fruit phenotype analysis

The 242-1 and 159-1 watermelon cultivars were used as maternal and paternal lines, respectively, to produce F_1 and F_2 progeny. The F_1 hybrid was produced by the crossing cultivars 242-1 and 159-1. The F_2 plants were subsequently produced by self-pollination of an F_1 plant and used for whole-genome sequencing analysis, genetic linkage mapping, QTL mapping, and phenotyping of fruit-related traits.

The β-carotene and lycopene of F_2 individuals were extracted from freeze-dried watermelon flesh according to the method described by Lee et al. (2022). Briefly, 0.1 g of finely crushed watermelon sample was added to 1 ml of ethanol containing 0.5 mM butylated hydroxytoluene and mixed carefully. Then, 3 ml of petroleum ether was added and vortexed, followed by 8 ml of 20% NaCl and additional vortexing for 1 min. After centrifugation at 3,000 rpm for 10 min, the supernatant was collected, and 1:1 (v/v) Na_2SO_4 was added and mixed. Finally, the sample was filtered with a 13-mm 0.2-μm Polytetrafluoroethylene (PTFE) Syringe Filter (Advantec Co., Tokyo, Japan) for analysis. Lycopene content was determined with an UltiMate 3000RS High performance liquid chromatography (HPLC) system (Thermo Fisher Scientific Inc.) equipped with a Kinetex 2.6-μm C18 100A reversed-phase column 100 × 4.60 mm (Phenomenex, Torrance, CA, USA), using 78% methanol for phase A and 100% ethyl acetate for phase B at a flow rate of 1 ml/min. The separation conditions were as follows: 0–8 min, 70% solution B; 8–10 min, 60% solution B; 10–12 min, 100% solution B; 12–14.01 min, 0% solution B; and 14.01–20 min, 100% solution B. Trans-lycopene (Sigma-Aldrich, St. Louis, MO, USA) and β-carotene (Sigma-Aldrich) were used as quantitative standards. Lycopene and β-carotene were quantified in the samples by measuring absorbance at 450 and 660 nm. The flesh color and rind color were expressed as red, green, and blue values using a Red-Green-Blue image analysis system (Aït-Aissa et al., 2018). Descriptive statistics and other statistical analyses were evaluated using SPSS v18.0 software (SPSS Inc., Chicago, IL, USA).

Whole-genome sequencing of F_2 populations and raw variant calling

For whole-genome resequencing, total genomic DNA was isolated from fresh young leaf tissues of F_2 plants and two parental plants (242-1 and 159-1) using the DNeasy Mini Kit (QIAGEN, Hilden, Germany). DNA samples were combined using different barcode adapters (5 μl each) and purified using the QIAquick PCR Purification Kit (cat. no. 28104, QIAGEN) according to the manufacturer's instructions. A 150-bp PE library with an insert size of 350 bp was constructed for each individual plant using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs). The qualified libraries were sequenced using an Illumina Hi Seq 2500 platform to produce 150-bp PE reads. Quality trimming was performed with the sliding window (4:20), average quality (20), and minimum read size (36) options in the Trimmomatic v 0.38 program. The trimmed reads were aligned to the 242-1 genome (female) using BWA-MEM aligner. After correcting the map file information using Picard and GATK codebases, a variant call was performed to generate SNP and InDel information for each sample. To select high-quality SNPs, variants with biallelic extraction, depth of coverage < 10, genotype quality < 30, or missing rate < 10% were removed. A total of 36,632 contigs were created for the 242-1 cultivar, and candidate InDel loci were compared to corresponding loci in the 159-1 cultivar. We selected InDel loci with a variant ratio > 0.8, InDel size > 30 bp, and number of reads > 20 for more accurate genotyping.

High-resolution mapping and QTL analysis of qualitative traits

A genetic map was constructed using the IciMapping v4.2 software (Institute of Crop Science Chinese Academy of Agricultural Sciences, Beijing, China) according to the following process: binning to remove duplicate markers, placing markers in linkage groups, ordering markers within linkage groups, and adjusting marker intervals within linkage groups (Meng et al., 2015). Markers were first filtered out by comparison of the offspring genotype distribution with the expected Mendelian proportions (1:2:1). A default value of data tolerance based on the X^2 test ($p < 0.05$) was used to discard highly separated markers. Then, markers were grouped using a maximum threshold recombination fraction value of 0.3. The recombination frequency was converted to map distance by Kosambi mapping (Kosambi, 2016). QTL analysis was performed using the IciMapping v4.2 software (Meng et al., 2015) with the inclusive composite interval mapping method. The logarithm of the odds (LOD) threshold was analyzed by the permutation method with 1,000 repeats to limit the type I error to 0.05 or less.

Results

Genome sequencing and assembly

We assembled and annotated a complete genome for Korean inbred watermelon cultivars 242-1 and 159-1 (Supplementary Figure S1). Nanopore long-read sequencing and Illumina short-read sequencing generated 7.7 Gb (22× coverage) and 8.6 Gb (25× coverage) long-read sequences and 15.6 Gb (44.6× coverage) and 14.1 Gb (40.3× coverage) high-quality trimmed short-read sequences for the 242-1 and 159-1 cultivars, respectively (Supplementary Table S1). The assembled long-read sequences of cultivars 242-1 and 159-1 had total lengths of 361.7 and 362.1 Mb, containing 43 and 103 initial contigs with N50 lengths of 16.0 and 6.3 Mb, respectively (Table 1). Pseudomolecules were constructed at the chromosome level based on the *C. lanatus* 97103 genome. Initial contigs were assembled *de novo* and clustered into 11 chromosomes ranging in length from 24.4 to 37.8 Mb for the 242-1 cultivar and 26.7 Mb to 37.9 Mb for the 159-1 cultivar (Supplementary Table S2). The lengths of most of the assembled chromosomes were very similar between 242-1 and 159-1, except that chromosome 4 of cultivar 242-1 was 24.4 Mb, which was about 2.7 Mb shorter than that of *C. lanatus* 97103 and about 2.3 Mb shorter than that of cultivar 159-1. We also used long-read sequencing to assemble the chloroplast genomes of cultivars 242-1 and 159-1, which provided new usable genetic information (Supplementary Figure S2).

We assessed the completeness of the genome assemblies by aligning published RNA-seq reads and performing BUSCO analysis. The full-length Illumina short-read transcripts of five watermelon tissues (flesh, SRR10389406; leaf, SRR3156561; light green rind, SRR10803496; root, SRR12079410; and shoot apex, SRR3156569) were aligned to the genome sequences of 242-1 and 159-1 (Wang et al., 2021). The results showed that more than 83.2% of the full-length transcripts could be mapped to the retained reads (Supplementary Table S3). A total of 1,548 (95.9%) and 1,547 (95.8%) BUSCOs were detected in the assembled genome of cultivars 242-1 and 159-1, respectively, indicating equivalent assembly quality compared to other assembled watermelon genomes (Figure 2C; Supplementary Table S4). The assembled genomes had similar chromosome synteny with the 97103 reference genome, which was expected given that the 242-1 and 159-1 pseudochromosomes were generated according to the genes and sequence of the 97103 genome. We found a high number of syntenic blocks in most of the chromosomes, except for chromosome 4. The syntenic blocks between the 242-1 assembly and the 97103 genome had different patterns in chromosome 4 (Supplementary Figure S3). Whole-genome nucleotide alignment showed that both assembled genomes aligned closely with the 97103 reference genome (Supplementary Figure S4, S5).

Repeat sequence annotation and gene prediction

We mapped the distributions of protein-coding genes, long terminal repeat (LTR)/Gypsy and LTR/Copia retrotransposons, and DNA transposons in the genome 242-1 and 159-1 to obtain an overview of the genome organization (Figures 2A, B). Analysis of repeat sequences revealed that the 242-1 and 159-1 assemblies contained 206 Mb (57.6%) and 207 Mb (57.5%) of repeat sequences and only 87 Mb (24.0%) and 86 Mb (24.2%) of assigned genes and gene-related sequences, respectively (Supplementary Figure S6). LTR retrotransposons were the most abundant repeat elements, comprising 46.4% and 45.9% of the 242-

TABLE 1 Comparison of the two assembled genomes (242-1 and 159-1) with two reference watermelon genomes (Charleston Gray and 97103).

Quality metric	Charleston Gray	97103	242-1	159-1
Total sequence length (bp)	397,829,775	365,450,462	361,732,851	362,131,170
Number of contigs	20,073	358	43	103
Contig N50 (bp)	48,885	2,312,425	16,071,105	6,343,009
Number of scaffolds	6,794	113	N/A	N/A
Scaffold N50 (bp)	33,982,080	35,099,344	N/A	N/A
Number of genes	22,546	23,440	23,921	24,451

N/A, Not applicable.

10.3389/fpls.2023.1142856

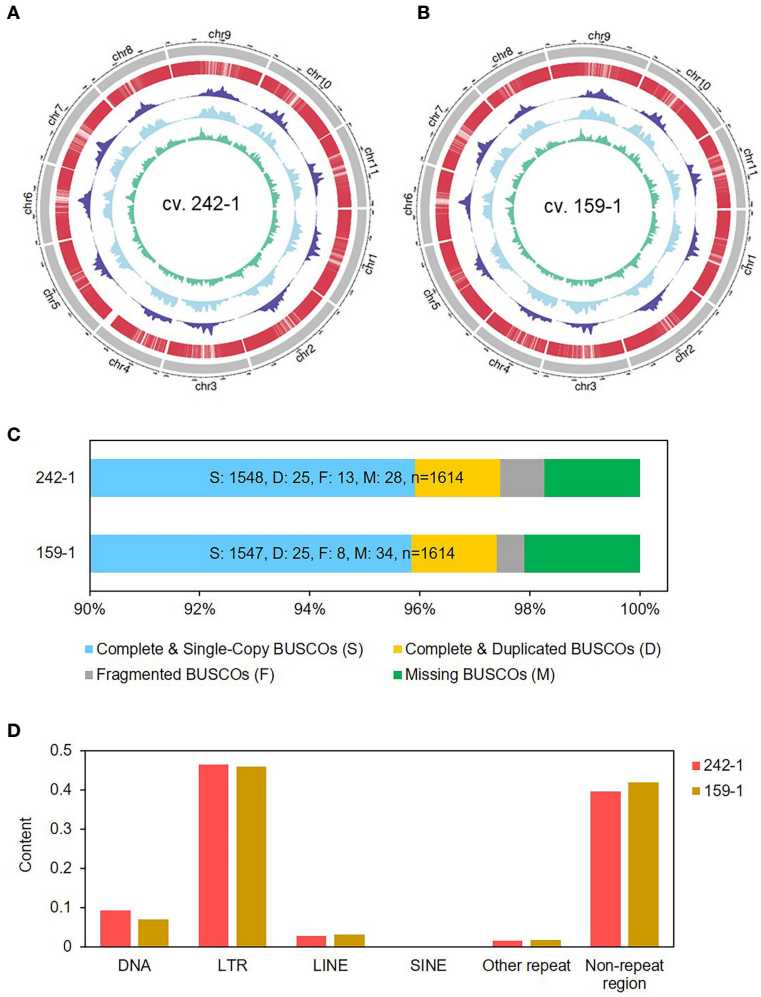


FIGURE 2
Genome characterization of the two inbred parental watermelon lines. **(A, B)** The outermost circle is the ideogram of 11 chromosomes in Mb scale, enclosing concentric circles of gene distribution (red), Gypsy content (navy blue), Copia content (sky blue), and DNA transposon content (green) in 1-Mb scale. **(C)** BUSCOs of the assembled watermelon genomes analyzed by using BUSCO v4.1.4 software. **(D)** Repetitive contents of the watermelon genomes analyzed by Repeat Masker with Repbase.

1 and 159-1 assemblies, respectively, followed by DNA transposons (9.3% and 7.1%), long interspersed nuclear elements (3.0% and 3.2%; Figure 2D; Supplementary Table S5). Gene annotations of the 242-1 and 159-1 assemblies predicted 23,921 and 24,451 protein-coding genes, respectively (Table 1). Of those, 21,358 (89.3%) in the 242-1 assembly and 21,321 (87.2%) in the 159-1 assembly have putative functional descriptions in public databases (Supplementary Table S6).

Carotenoid biosynthesis-related gene analysis

We compared the expression of protein-coding genes involved in carotenoid biosynthesis between the 242-1 and 159-1 cultivars. Comparison of the Illumina short-read transcripts to annotated sequences in the KEGG database identified 24 genes in the carotenoid biosynthetic pathway that were expressed in 242-1 and 159-1: three phytoene synthases (*PSY1*, *PSY2*, and *PSY3*), three phytoene desaturases (*PDS1*, *PDS2*, and *PDS3*), one ζ-carotene isomerase (*Z-ISO*), one ζ-carotene desaturase (*ZDS*), three β-carotene isomerases (*CRTISO1*, *CRTISO2*, and *CRTISO3*), four lycopene cyclases (*LCYB1*, *LCYB2*, *LCYB/E*, and *LCYE*), two β-carotene hydroxylases (*CHYB1* and *CHYB2*), five zeaxanthin epoxidases (*ZEP1*, *ZEP2*, *ZEP3*, *ZEP4*, and *ZEP5*), and two violaxanthin de-epoxidases (*VDE1* and *VDE2*) (Figure 3; Supplementary Table S7). A heatmap of these genes based on FPKM values calculated from the Illumina short-read transcripts showed that upstream genes for phytoene synthesis (*PSY1*, *PSY2*, and *PSY3*) and the downstream genes for zeaxanthin epoxide synthesis (*ZEP1*, *ZEP2*, *ZEP3*, *ZEP4*, *ZEP5*, *VDE1*, and *VDE2*) in the carotenoid biosynthesis pathway had similar expression in the 242-1 and 159-1 cultivars. By contrast, most of the genes for lycopene synthesis (*PDS2*, *ZDS*, and *CRTISO1*) and carotene synthesis (*LCYB1*, *LCYB/E*, and *CHYB1*) had different expressions between 242-1 and 159-1.

10.3389/fpls.2023.1142856

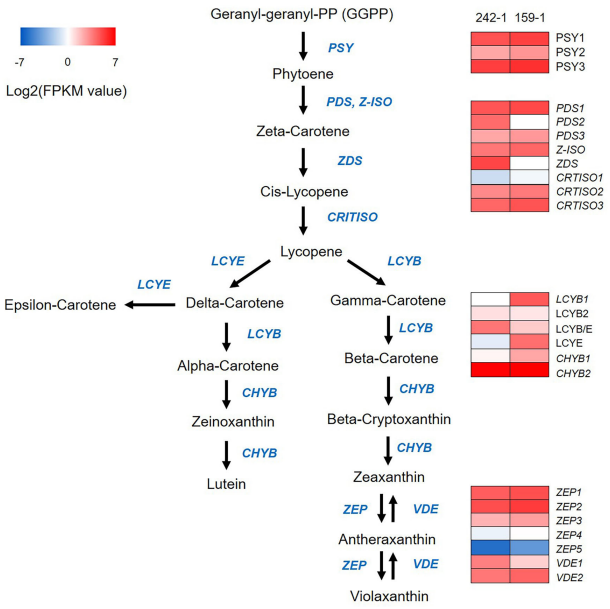


FIGURE 3
A heatmap diagram of expression levels of carotenoid biosynthetic pathway genes identified by KEGG analysis. The heatmap was drawn according to FPKM values calculated from Illumina short-read transcripts. Columns and rows in the heatmap represent samples and genes, respectively. Sample names are displayed above the heatmap. Color scale indicates fold changes in gene expression. PSY, phytoene synthase; PDS, phytoene desaturase; Z-ISO, zeta-carotene isomerase; CRTISO, beta-carotene isomerase; LCYB, lycopene beta-cyclase; LCYB/E, lycopene beta/epsilon cyclase; LCYE, lycopene epsilon cyclase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase.

Genotyping and genetic linkage mapping

For genetic linkage mapping, a total of 1.1 Tb of WGS data comprising 7.5 billion reads were generated from the two parental samples (242-1 and 159-1) and the 87 F_2 samples (Supplementary Table S8). A total of 2,029,598 raw SNPs and 869,298 InDels were identified among all individuals using VCF tools (Danecek et al., 2011). Of those, 686,357 SNPs and 937 InDels passed through quality filtering (see Materials and Methods; Supplementary Table S9). We selected 140,650 homozygous SNPs by screening for SNPs with the same genotype as the 242-1 (maternal) sequence (AA) and a different genotype than the 159-1 (paternal) sequence (BB). After the linkage maps were subjected to a binning process to remove duplicate markers, a total of 2,319 SNP markers were selected. In addition, a total of 167 InDels were selected and used to genotype the individual samples. These InDel markers were tested by PCR and electrophoresis analysis to confirm the genotypes of 242-1, 159-1, and F_1 samples. As a result, 126 InDel markers were finally selected for genotyping of the F_2 samples (Supplementary Table S10). On the basis of the electrophoresis bands of the selected InDel markers, each F_2 individual was genotyped for each marker as maternal type, paternal type, or F_1 type (heterozygous) (Supplementary Figure S7).

A total of 780 (33.6%) SNP markers and 45 (35.7%) InDel markers did not conform to a Mendelian ratio of 1:2:1 in the F_2 samples and were discarded. As a result, a total of 1,619 SNP and InDel markers were identified and prepared for genetic linkage mapping (Supplementary Tables S11; Table S12). These markers were sorted into 11 linkage groups (Figure 4), which is consistent with the reported haploid chromosome number of watermelon (Guo et al., 2013). The genetic linkage map spanned 3,036.9 cM, with an average interval of 1.87 cM. The genetic length of each chromosome ranged from 187.37 cM (chromosome 11) to 379.57 cM (chromosome 01), with an average SNP distance of 1.56–2.48 cM (Figure 4 and Supplementary Table S11).

QTL mapping of fruit quality-related traits

Descriptive statistics for fruit quality-related traits are shown in Table 2; Supplementary Table S13. As the main components of flesh pigment, individuals in the F_2 mapping family had lycopene levels ranging from 0.13 to 38.3 μg/g and β-carotene levels ranging from 0.17 to 5.24 μg/g. The flesh and rind color of the red, green, and blue values were obvious differences among F_2 individuals. We performed Kolmogorov-Smirnov and Shapiro-Wilk tests (Razali and Wah, 2011) to investigate the normality of each fruit quality-related trait among the samples. The frequency distributions of the traits are shown in Supplementary Figure S8. Traits related to red or green color in the flesh or rind were strongly correlated with each other ($r = 0.92$ for flesh color, $r = 0.99$ for rind color, P-value < 0.01 for all). Traits related to lycopene content were also strongly correlated with traits related to red or green flesh color ($r = -0.8$ for red flesh, $r = -0.84$ for green flesh, P-value < 0.01 for all; Supplementary Figure S9). This result showed that the red or green flesh color were closely related to lycopene content in watermelon.

Fifteen QTL intervals (one for fruit shape index, one for skin thickness, two for lycopene content, two for β-carotene content,

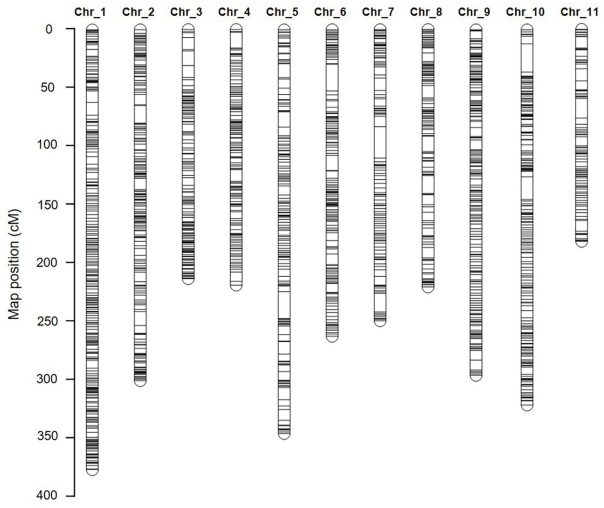


FIGURE 4
The high-density genetic linkage map of watermelon. The inter-specific linkage map of the inbred watermelon lines 242-1 and 159-1 harboring 1,619 loci. SNP and InDel markers are represented in black.

10.3389/fpls.2023.1142856

TABLE 2 Descriptive statistics of the traits investigated in the F_2 population.

Trait	Mean ± SE	5% Trimmed mean	Median	Variance	Standard deviation	Minimum	Maximum	IQR	Skewness	Kurtosis
FSI	1.23 ± 0.02	1.2	1.13	0.06	0.24	0.94	2.0	0.24	1.41	1.34
SKT	0.92 ± 0.03	0.91	0.9	0.07	0.27	0.5	1.6	0.4	0.54	−0.26
L	8.80 ± 1.02	7.89	4.01	89.82	9.48	0.1	38.3	15.1	1.18	0.65
BC	2.02 ± 0.12	1.95	1.96	1.2	1.1	0.17	5.24	1.35	0.89	0.82
FR	220.09 ± 2.20	220.63	223	421.4	20.5	180	250	39	−0.33	−1.2
FG	139.02 ± 6.66	140.14	155	3860.9	62.1	27	225	118	−0.34	−1.38
FB	76.90 ± 3.85	75.73	74	1290	35.9	4	173	53	0.49	−0.24
RR	160.02 ± 9.41	162.62	226	7706.1	87.8	28	244	179	−0.46	−1.71
RG	141.24 ± 7.14	153.47	193	4411.5	66.4	36	230	125	−0.56	−1.45
RB	70.13 ± 5.00	68.61	60	2171.9	46.6	1	173	72	0.5	−0.8

IQR, interquartile range; FSI, fruit shape index; SKT, skin thickness (cm); L, lycopene (μg/g); BC, β-carotene (μg/g); FG, flesh color (green); FR, flesh color (red); RB, rind color (blue); RG, rind color (green); RR, skin color (red).

four for flesh color, and five for skin color) associated with fruit quality-related traits were identified on chromosomes 1, 2, 3, 4, 6, and 8 by permutation tests ($p < 0.05$, LOD > 4.3; Figure 5; Table 3). The QTL interval with the highest LOD value of 25.4 was located at 171 cM between the markers ClaB_Chr04_14736633 and ClaB_Chr04_14817220 and accounted for 47.3% of the percentage variance explained (PVE). Eight QTL intervals contributed simultaneously to two traits: *Cqly2.*1 and *Cqbc2.1* to lycopene and β-carotene levels; *Cqfg2.1*, *Cqfr2.1*, *Cqfg4.1*, and *Cqfr4.1* to green and red fruit colors; and *Cqrg4.1* and *Cqrr4.1* to read and green skin colors.

We used BLAST to search for gene sequences in the 15 fruit quality-related QTL intervals based on the annotation of the 97103 reference genome in the Cucurbit Genomics Database (http://cucurbitgenomics.org/blast) (Zheng et al., 2019). We identified a total of 302 candidate genes in the 15 QTL intervals. The QTL interval with the most candidate genes was *Cqrb6.1*, which contained 67 candidate genes, followed by *Cqrb1.1* with 66 candidate genes, whereas *Cqskt2.1*, *Cqrg4.1*, and *Cqrr4.1* each contained one candidate gene (Supplementary Table S14). We compared the candidate genes between the parental cultivars using the resequencing data and identified 33 genes that had variation in an exon (Supplementary Table S15). We then performed gene ontology-enrichment analysis and KEGG pathway analysis to investigate the functions of these 33 genes. The results identified two genes involved in carotenoid biosynthesis that were also linked to flesh color, phytoene synthase 1 (*PSY1*, *Cla97C01G008760*), and red chlorophyll catabolite reductase (*RCCR*, *Cla97C02G038270*), which were located in the QTL intervals *Cqbc1.1* and *Cqly2.1*, respectively (Supplementary Table S15). *RCCR* had a single-nucleotide variant (C → G) in the first exon at the 25,790,540 bp of chromosome 2 that resulted in conversion of an asparagine to a lysine. *PSY1* had a single-nucleotide variant (A → G) in the first exon at 9,539,129 bp of chromosome 1 that resulted in conversion of a lysine to glutamic acid (Figure 6). In 242-1 and 159-1, the *PSY1* variant encoding the glutamic acid was exclusive to 242-1, which was the only cultivar with non-red flesh (orange flesh).

Discussion

The three previously assembled domesticated watermelon genomes (Charleston Gray, 97103, and G42) are all from cultivars with green skin and red flesh (Guo et al., 2013; Wu et al., 2019; Deng et al., 2022). Our genome assemblies provide new information on watermelon cultivars with black or yellow skin and orange flesh. Chromosome 4 of the red-flesh, yellow-rind 159-1 cultivar is similar to that of the red-flesh 97103 reference cultivar, whereas chromosome 4 of the orange-flesh, black-rind 242-1 cultivar has a different construction. There are many genes and regions associated with fruit color (Guo et al., 2019; Wang et al., 2019; Wang et al., 2021) and rind color (Park et al., 2016; Dou et al., 2018; Liu et al., 2020) on chromosome 4, which indicates the specificity of the 242-1 genome. The distinguishing characteristics of the 242-1 and 159-1 watermelon genomes are mainly reflected in flesh and rind colors. The carotenoid biosynthesis-related genes identified in the two new assembled genomes provide a basis for the mapping of watermelon fruit color and rind color.

The color of watermelon flesh is mainly linked to the composition of xanthophylls and carotenes (mainly phytoene, lycopene, and β-carotene) (Wang et al., 2021). We found that two lycopene biosynthesis-related genes, *PDS2* and *ZDS*, had higher expression in the orange-fleshed 242-1 cultivar than in the red-

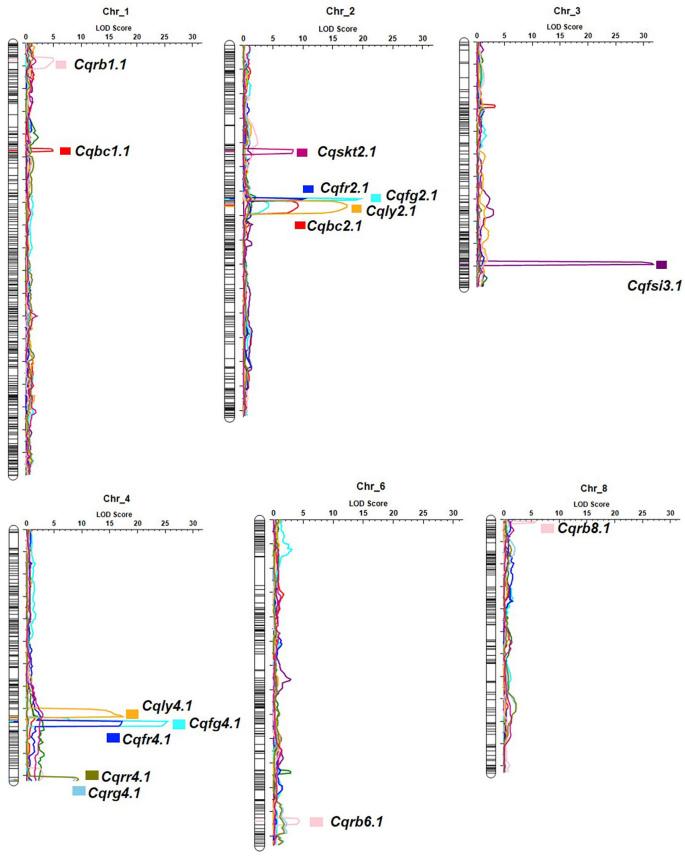


FIGURE 5
Genome-wide QTL mapping of fruit quality-related traits in the F_2 population. The boxes on the right of the chromosomes represent QTLs for fruit shape index (Cqfsi3.1), skin thickness (Cqskt2.1), lycopene (Cqly2.1 and Cqly4.1), β-carotene (Cqbc1.1 and Cqbc2.1), flesh color (Cqfg2.1, Cqfg4.1, Cqfr2.1, and Cqfr4.1), and rind color (Cqrb1.1, Cqrb6.1, Cqrb8.1, Cqrg4.1, and Cqrr4.1).

fleshed 159-1 cultivar. The *PDS*, *ZDS*, and *CRTISO* genes are considered to play only connecting roles in lycopene biosynthesis and accumulation (Kato et al., 2004). These results suggest that *PDS2* catalysis of ζ-carotene biosynthesis and *ZDS* catalysis of cis-lycopene biosynthesis are more active in orange-fleshed watermelon (242-1) than in red-fleshed watermelon (159-1), but this difference does not result in a change of lycopene content. The expression of lycopene cyclase genes (*LCYB1*, *LCYB/E*, and *LCYE*) that regulate α-carotene, β-carotene, γ-carotene, δ-carotene, and ϵ-carotene biosynthesis from lycopene was also different between the 242-1 and 159-1 cultivars, suggesting that the two watermelon cultivars have very different mechanisms of carotene synthesis and accumulation.

A single-nucleotide variant of the *PSY1* (*Cla97C01G008760*) gene located in the β-carotene–mapping QTL interval *Cqbc1.1* differs between the orange-fleshed cultivar 242-1 and the three red-fleshed cultivars 97103, Charleston Gray, and 159-1, suggesting that this variant has a strong impact on flesh color. Phytoene synthase is the first enzyme in the carotenoid biosynthesis pathway and converts geranylgeranyl diphosphate into phytoene. This enzyme defines the size of the carotenoid pool, which suggests that the variation that we detected in *PSY1* affects the carotenoid composition of the watermelon fruit, resulting in the orange color in cultivar 242-1. Orange-flesh watermelon is a landrace of red-flesh watermelon, and coral-red flesh (*Y*) is dominant to orange flesh (y^o) (Bang et al., 2010; Guo et al., 2019). This implies that red-flesh watermelon evolved from orange-flesh watermelon during a process of selective breeding and domestication, during which variation of *PSY1* arose and resulted in a decrease in β-carotene content and an increase in lycopene content. On the basis of QTL mapping of variants related to lycopene and β-carotene synthesis, we found an *RCCR* gene in the QTL intervals *Cqly2.1* and *Cqbc2.1*. RCCR is essential in chlorophyll degradation during leaf senescence and fruit ripening in higher plants (Hörtensteiner, 2006), and this chlorophyll degradation is closely linked to the formation of carotenoids (Meier et al., 2011; Tripathy and Pattanayak, 2012; Leng et al., 2017). Our results suggest that the *RCCR* gene participates in the regulation of lycopene and β-carotene biosynthesis and accumulation in watermelon fruit. A previous study reported that levels of various carotenoids, such as lutein, β-carotene, violaxanthin, and zeaxanthin, were highly increased in *RCCR*-silenced tobacco plants (Dong et al., 2022). Furthermore, although lycopene was not mentioned in that report, expression of the phytoene synthase *NbPSY1* and the β-carotene isomerase *NbCRTISO*, both key genes in lycopene synthesis, was also increased after *RCCR* silencing (Dong et al., 2022). This may be

TABLE 3 Characteristics of the fruit quality-related QTLs in the watermelon (LOD > 4.3).

QTL	Trait	Chr	Position (cM)	Marker interval	LOD	PVE (%)	Add	Dom	Left CI	Right CI
Cqfsi3.1	FSI	3	198	ClaB_Chr03_28459812–ClaB_Chr03_28508794	31.7	79.6	0.3	−0.17	196.5	198.5
Cqskt2.1	SKT	2	87	ClaB_Chr02_30520465–ClaB_Chr02_30520917	8.3	34.9	−0.2	0.0	85.5	89.5
Cqly2.1	L	2	132	ClaB_Chr02_26461833–ClaB_Chr02_25758590	17.3	36.0	−9.6	−6.6	129.5	135.5
Cqly4.1	L	4	167	ClaB_Chr04_16911823–Clg-InDel-165	17.4	32.5	−8.6	−6.2	165.5	167.5
Cqbc1.1	BC	1	94	ClaB_Chr01_9498243– ClaB_Chr01_9770736	4.9	15.7	59.7	−0.1	92.5	95.5
Cqbc2.1	BC	2	131	ClaB_Chr02_26461833–ClaB_Chr02_25758590	9.2	34.0	−86.9	−24.0	128.5	136.5
Cqfg2.1	FG	2	126	ClaB_Chr02_26728528–ClaB_Chr02_26706701	19.5	30.6	51.3	−0.4	125.5	127.5
Cqfg4.1	FG	4	171	ClaB_Chr04_14817220–ClaB_Chr04_14736633	25.4	47.3	64.3	−3.2	170.5	174.5
Cqfr2.1	FR	2	126	ClaB_Chr02_26728528–ClaB_Chr02_26706701	10.2	23.7	13.7	4.3	125.5	127.5
Cqfr4.1	FR	4	171	ClaB_Chr04_14817220–ClaB_Chr04_14736633	17.2	47.9	20.0	6.4	170.5	175.5
Cqrb1.1	RB	1	15	ClaB_Chr01_1527902–ClaB_Chr01_2118855	5.0	14.9	−4.2	−35.4	12.5	19.5
Cqrb6.1	RB	6	247	ClaB_Chr06_26810713–ClaB_Chr06_27345836	4.4	12.4	−7.2	30.7	244.5	250.5
Cqrb8.1	RB	8	2	ClaB_Chr08_27454491–ClaB_Chr08_27488871	5.7	17.0	−28.1	3.9	1.5	3.5
Cqrg4.1	RG	4	223	ClaB_Chr04_6127109–ClaB_Chr04_6095999	8.7	35.3	−54.7	21.1	220.5	224.0
Cqrr4.1	RR	4	223	ClaB_Chr04_6127109–ClaB_Chr04_6095999	9.3	37.3	−73.8	30.2	220.5	224.0

PVE, percentage variance explained; Add, QTL with additive effect only; Dom, QTL with complete dominance; CI, confidence interval (cM); FSI, fruit shape index; SKT, skin thickness; L, lycopene; BC, β-carotene; FG, flesh color (green); FR, flesh color (red); RB, rind color (blue); RG, rind color (green); RR, skin color (red).

because there is no important mechanism for lycopene accumulation in tobacco leaves. In tomatoes, the lycopene and chlorophyll contents are inversely proportional, and lycopene synthesis and chlorophyll decomposition are synchronized (Carrillo-López and Yahia, 2014; Arthanari and Dhanapalan, 2019). Our results confirmed that *RCCR* plays an important role in controlling the chlorophyll and carotenoid contents in leaves and fruits of higher plants, although the mechanism varies among plants.

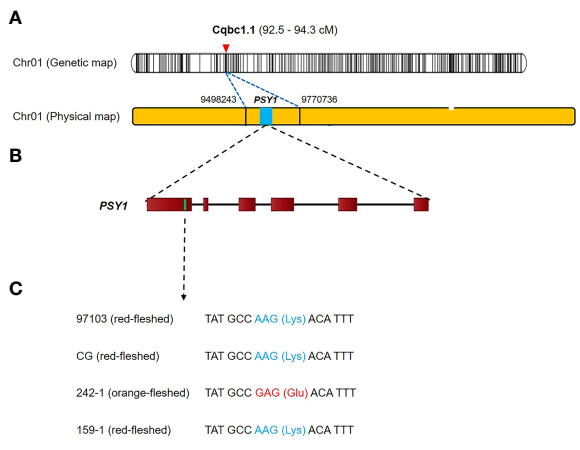


FIGURE 6
Genetic mapping and phenotypic analysis of the carotenoid biosynthesis gene *PSY1* in watermelon. **(A)** The preliminary genetic map and physical map. The *PSY1* gene is located on chromosome 1. **(B)** Structure of the *PSY1* gene. A nonsynonymous SNP mutant is located in the first exon at 9,539,129 bp on chromosome 1. **(C)** The SNP (A → G) results in the conversion of a lysine (Lys) to a glutamic acid (Glu).

Like the flesh, the rind of watermelon fruit can have various colors, including black, dark green, light green, or yellow (Guo et al., 2013). Yellow-rind watermelon also has high carotenoid contents (Dou et al., 2018); however, there have been few genetic studies of the mechanism of rind color inheritance in watermelons. Previously, rind color-related loci were detected at the ends of chromosomes 4, 6, 8, and 10 (Jones et al, 2014; Park et al., 2016; Park et al., 2018). Genes involved in fruit skin color were previously linked to pigment binding; chloroplast membrane development; anthocyanin, porphyrin, and chlorophyll metabolism; hormone signal transduction; photosynthesis; and carotenoid biosynthesis (Hu et al, 2020; Liu et al., 2020). We found five rind color-related QTL intervals (Cqrb1.1, Cqrg4.1, Cqrr4.1, Cqrb6.1, and Cqrb8.1) on chromosomes 1, 4, 6, and 8, suggesting that rind color-related loci are located mainly on chromosomes 4, 6, and 8. We identified five candidate genes with CDS variation that might be involved in rind color (Cla97C04G070840, Cla97C06G124570, Cla97C06G124600, Cla97C06G124910, and Cla97C08G160580). In previous studies, the Dgo gene (Cla97C04G068530/Cla002769), which is involved in chlorophyll synthesis, was found to be involved in determining the background rind color (Park et al., 2016), and the ClCGMenG gene, which encodes 2-phytyl-1,4-beta-naphthoquinone methyltransferase, was found to be involved in determining dark green or light green color in the rind (Li et al., 2019). However, neither of these genes mapped to our candidate regions. Cultivars 242-1 and 159-1 have black and yellow rinds, respectively, which is different from the green-rind watermelons used in previous

studies (Li et al., 2009; Park et al., 2016; Park et al., 2018; Li et al., 2019). The black and yellow rind colors may be the result of loci and candidate genes identified in our screen.

In summary, we performed genetic mapping of traits related to watermelon fruit quality, especially flesh color and rind color, and identified important loci and genes. A phytoene synthase encoded by *Cla97C01G008760* is likely to be a core element regulating the carotenoid metabolic pathways in the fruits of watermelon cultivars 242-1 and 159-1, which might have a chain effect on the downstream synthesis or accumulation of other substances, thereby changing the flesh color. We will continue to pay attention to this variation and use further genetic analysis to study the mechanisms that determine the color of watermelon flesh.

Data availability statement

The data presented in the study are deposited in the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) repository, accession number PRJNA924512 and PRJNA924516.

Author contributions

S-YK and YL designed the project. SN and SP collected and grew the plant material. SLe, SLi, ML, and JK performed the experiments. MK performed the data analysis. HN conducted software analysis and wrote the original draft. S-TK, A-YS, YL, and S-YK revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Korea Post-Genome Project of the Rural Development Administration (No. PJ013344012019), Korea, and the Korean Research Institute of Bioscience and Biotechnology Initiative Program.

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.2023.1142856/full#supplementary-material

References

Aït-Aissa, A., Zaddem, M., and Aïder, M. (2018). Red–Green–Blue (RGB) colour system approach to study the segregation and percolation in a mixture of white wheat flour and bleached wheat bran. *Int. J. Food Sci. Technol.* 53, 254–261. doi: 10.1111/ijfs.13581

Arthanari, M., and Dhanapalan, S. (2019). Quantification of β-carotene, lycopene, and chlorophyll content in tomato fruits of enrichment of chicken feathers composting. *Int. J. Recycl. Org. Waste Agric.* 8, 473–477. doi: 10.1007/s40093-019-0258-6

Bae, S. J., Islam, M. M., Kim, H. Y., and Lim, K. B. (2020). Induction of tetraploidy in watermelon with oryzalin treatments. *Hortic. Sci. Technol.* 38, 385–393. doi: 10.7235/HORT.20200037

Bang, H., Davis, A. R., Kim, S., Leskovar, D. I., and King, S. R. (2010). Flesh color inheritance and gene interactions among canary yellow, pale yellow, and red watermelon. *J. Am. Soc Hortic. Sci.* 135, 362–368. doi: 10.21273/jashs.135.4.362

Carrillo-López, A., and Yahia, E. M. (2014). Changes in color-related compounds in tomato fruit exocarp and mesocarp during ripening using HPLC-APcI+-mass spectrometry. *J. Food Sci. Technol.* 51, 2720–2726. doi: 10.1007/s13197-012-0782-0

Chomicki, G., Schaefer, H., and Renner, S. S. (2020). Origin and domestication of cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytol.* 226, 1240–1255. doi: 10.1111/nph.16015

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

Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol. Plant* 15, 1268–1284. doi: 10.1016/j.molp.2022.06.010

Dong, C., Zhang, M., Wei, F., Guo, Y., Qin, L., Wei, P., et al. (2022). Inhibition of red chlorophyll catabolite reductase improved chlorophyll and carotenoid synthesis in tobacco. *Plant Cell. Tissue Organ Cult.* 148, 687–698. doi: 10.1007/s11240-022-02228-1

Dou, J., Lu, X., Ali, A., Zhao, S., Zhang, L., He, N., et al. (2018). Genetic mapping reveals a marker for yellow skin in watermelon (*Citrullus lanatus* l.). *PloS One* 13, 1–15. doi: 10.1371/journal.pone.0200617

Dubey, S., Rajput, H., and Batta, K. (2021). Utilization of watermelon rind (*Citrullus lanatus*) in various food preparations: A review. *J. Agric. Sci. Food Res.* 12, 5–7. doi: 10.37273/chesci.cs205205361

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117

Guo, S., Sun, H., Zhang, H., Liu, J., Ren, Y., Gong, G., et al. (2015). Comparative transcriptome analysis of cultivated and wild watermelon during fruit development. *PloS One* 10, 1–21. doi: 10.1371/journal.pone.0130267

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470

Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., et al. (2003). Improving the arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi: 10.1093/nar/gkg770

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, 1–22. doi: 10.1186/gb-2008-9-1-r7

Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. *Annu. Rev. Plant Biol.* 57, 55–77. doi: 10.1146/annurev.arplant.57.032905.105212

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36, 2253–2255. doi: 10.1093/bioinformatics/btz891

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kato, M., Ikoma, Y., Matsumoto, H., Sugiura, M., Hyodo, H., and Yano, M. (2004). Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol.* 134, 824–837. doi: 10.1104/pp.103.031104

Kosambi, D. D. (2016). *The estimation of map distances from recombination values*. Eds. D. D. Kosambi and R. Ramaswamy (New Delhi Springer), 125–130. Available at: https://link.springer.com/chapter/10.1007/978-81-322-3676-4_16.

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 1–13. doi: 10.1186/s13059-019-1910-1

Lee, S., Park, G., Choi, Y., Park, S., Kim, H., Lee, O., et al. (2022). Whole-genome resequencing of near-isogenic lines reveals a genomic region associated with high trans-lycopene contents in watermelon. *Plants* 11, 1–14. doi: 10.3390/plants11010008

Leng, X., Wang, P., Wang, C., Zhu, X., Li, X., Li, H., et al. (2017). Genome-wide identification and characterization of genes involved in carotenoid metabolic in three stages of grapevine fruit development. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-04004-0

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

Li, B., Zhao, S., Dou, J., Ali, A., Gebremeskel, H., Gao, L., et al. (2019). Genetic mapping and development of molecular markers for a candidate gene locus controlling rind color in watermelon. *Theor. Appl. Genet.* 132, 2741–2753. doi: 10.1007/s00122-019-03384-3

Liu, D., Yang, H., Yuan, Y., Zhu, H., Zhang, M., Wei, X., et al. (2020). Comparative transcriptome analysis provides insights into yellow rind formation and preliminary mapping of the clyr (Yellow rind) gene in watermelon. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00192

Liu, C., Zhang, H., Dai, Z., Liu, X., Liu, Y., Deng, X., et al. (2012). Volatile chemical and carotenoid profiles in watermelons [Citrullus vulgaris (Thunb.) schrad (Cucurbitaceae)] with different flesh colors. *Food Sci. Biotechnol.* 21, 531–541. doi: 10.1007/s10068-012-0068-3

Luo, X., Chen, S., and Zhang, Y. (2022). PlantRep: a database of plant repetitive elements. *Plant Cell Rep.* 41, 1163–1166. doi: 10.1007/s00299-021-02817-y

Ma, G., Zhang, L., Kudaka, R., Inaba, H., Furuya, T., Kitamura, M., et al. (2021). Exogenous application of aba and naa alleviates the delayed coloring caused by puffing inhibitor in citrus fruit. *Cells* 10, 1–13. doi: 10.3390/cells10020308

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PloS Comput. Biol.* 14, 1–14. doi: 10.1371/journal.pcbi.1005944

Meier, S., Tzfadia, O., Vallabhaneni, R., Gehring, C., and Wurtzel, E. T. (2011). A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in arabidopsis thaliana. *BMC Syst. Biol.* 5, 1–19. doi: 10.1186/1752-0509-5-77

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

Nadeem, M., Navida, M., Ameer, K., Iqbal, A., Malik, F., Nadeem, M. A., et al. (2022). A comprehensive review on the watermelon phytochemical profile and their bioactive and therapeutic effects. *Korean J. Food Preserv.* 29, 546–576. doi: 10.11002/kjfp.2022.29.4.546

Paris, H. S. (2015). Origin and emergence of the sweet dessert watermelon, citrullus lanatus. *Ann. Bot.* 116, 133–148. doi: 10.1093/aob/mcv077

Park, G., Kim, J., Jin, B., Yang, H. B., Park, S. W., Kang, S. C., et al. (2018). Genome-wide sequence variation in watermelon inbred lines and its implication for marker-assisted breeding. *Hortic. Sci. Technol.* 36, 280–291. doi: 10.12972/kjhst.20180028

Park, S., Kim, K. T., Kang, S. C., and Yang, H. B. (2016). Rapid and practical molecular marker development for rind traits in watermelon. *Hortic. Environ. Biotechnol.* 57, 385–391. doi: 10.1007/s13580-016-0005-0

Perkins-Veazie, P., Collins, J. K., Davis, A. R., and Roberts, W. (2006). Carotenoid content of 50 watermelon cultivars. *J. Agric. Food Chem.* 54, 2593–2597. doi: 10.1021/jf052066p

Perkins-Veazie, P., Davis, A., and Collins, J. K. (2012). Watermelon: From dessert to functional food. *Isr. J. Plant Sci.* 60, 395–402. doi: 10.1560/IJPS.60.1.402

Porcher, M. H., Noriko, A., Mundy, A., Marinos Drakopoulos, E. M., Rocha, A., Li, J., et al. (2013) *Sorting citrullus names. multiling. multiscript plant name database*. Available at: https://www.plantnames.unimelb.edu.au/Sorting/Citrullus.html#lanatus-vulgaris-gr.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Razali, N. M., and Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal.* 2, 21–33.

Renner, S. S., Wu, S., Pérez-Escobar, O. A., Silber, M. V., Fei, Z., and Chomicki, G. (2021). A chromosome-level genome of a kordofan melon illuminates the origin of domesticated watermelons. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–9. doi: 10.1073/pnas.2101486118

Rhee, S. J., Han, B. K., Jang, Y. J., Sim, T. Y., and Lee, G. P. (2015). Construction of a genetic linkage map using a frame set of simple sequence repeat and high-resolution melting markers for watermelon (Citrullus spp.). *Hortic. Environ. Biotechnol.* 56, 669–676. doi: 10.1007/s13580-015-0110-5

Roach, M. J., Schmidt, S., and Borneman, A. R. (2018). Purge haplotigs: Synteny reduction for third-gen diploid genome assemblies. *BMC Bioinf.* 19, 1–10. doi: 10.1186/s12859-018-2485-7

Sandlin, K., Prothro, J., Heesacker, A., Khalilian, N., Okashah, R., Xiang, W., et al. (2012). Comparative mapping in watermelon [Citrullus lanatus (Thunb.) matsum. et nakai]. *Theor. Appl. Genet.* 125, 1603–1618. doi: 10.1007/s00122-012-1938-z

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Simpson, K., Quiroz, L. F., Rodriguez-Concepción, M., and Stange, C. R. (2016). Differential contribution of the first two enzymes of the MEP pathway to the supply of metabolic precursors for carotenoid and chlorophyll biosynthesis in carrot (Daucus carota). *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01344

Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6, 1–11. doi: 10.1186/1471-2105-6-31

Sun, L., Zhang, Y., Cui, H., Zhang, L., Sha, T., Wang, C., et al. (2020). Linkage mapping and comparative transcriptome analysis of firmness in watermelon (Citrullus lanatus). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00831

Tadmor, Y., King, S., Levi, A., Davis, A., Meir, A., Wasserman, B., et al. (2005). Comparative fruit colouration in watermelon and tomato. *Food Res. Int.* 38, 837–841. doi: 10.1016/j.foodres.2004.07.011

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 4, 1–14. doi: 10.1002/0471250953.bi0410s25

Wang, C.-n., Luan, F.-s., Liu, H.-y., Davis, A. R., Zhang, Q., Dai, Z.-y., et al. (2021). Mapping and predicting a candidate gene for flesh color in watermelon. *J. Integr. Agric.* 20, 2100–2111. doi: 10.1016/S2095-3119(20)63487-6

Wang, C., Qiao, A., Fang, X., Sun, L., Gao, P., Davis, A. R., et al. (2019). Fine mapping of lycopene content and flesh color related gene and development of molecular marker–assisted selection for flesh color in watermelon (Citrullus lanatus). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01240

Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L., et al. (2019). Genome of 'Charleston gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. national plant germplasm system watermelon collection. *Plant Biotechnol. J.* 17, 2246–2258. doi: 10.1111/pbi.13136

Xiao, J., Chen, S.-y., Sun, Y., Yang, S.-d., and He, Y. (2022). Differences of rhizospheric and endophytic bacteria are recruited by different watermelon phenotypes relating to rind colors formation. *Sci. Rep.* 12, 1–13. doi: 10.1038/s41598-022-10533-0

Zerpa-Catanho, D., Zhang, X., Song, J., Hernandez, A. G., and Ming, R. (2021). Ultra-long DNA molecule isolation from plant nuclei for ultra-long read genome sequencing. *STAR Protoc. 2*, 1–17. doi: 10.1016/j.xpro.2021.100343

Zhao, W., Lv, P., and Gu, H. (2013). Studies on carotenoids in watermelon flesh. *Agric. Sci.* 04, 13–20. doi: 10.4236/as.2013.47a003

Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2019). Cucurbit genomics database (CuGenDB): A central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47, D1128–D1136. doi: 10.1093/nar/gky944

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fpls.2023.1138415

OPEN ACCESS

EDITED BY
Qiusheng Kong,
Huazhong Agricultural University, China

REVIEWED BY
Lei Zhang,
Jiangsu Normal University, China
Tangren Cheng,
Beijing Forestry University, China

*CORRESPONDENCE
Li Yuan
lyuan@nwafu.edu.cn
Ruimin Zhang
zrm0923@sdau.edu.cm
Chunhua Wei
xjwend020405@nwsuaf.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 05 January 2023
ACCEPTED 08 February 2023
PUBLISHED 02 March 2023

CITATION
Yue Z, Pan X, Li J, Si F, Yin L, Hou Y, Chen X, Li X, Zhang Y, Ma J, Yang J, Li H, Luan F, Huang W, Zhang X, Yuan L, Zhang R and Wei C (2023) Whole-transcriptome analyses identify key differentially expressed mRNAs, lncRNAs, and miRNAs associated with male sterility in watermelon. *Front. Plant Sci.* 14:1138415. doi: 10.3389/fpls.2023.1138415

Whole-transcriptome analyses identify key differentially expressed mRNAs, lncRNAs, and miRNAs associated with male sterility in watermelon

Zhen Yue[1], Xiaona Pan[1], Jiayue Li[1], Fengfei Si[1], Lijuan Yin[1], Yinjie Hou[1], Xiaoyao Chen[1], Xin Li[1], Yong Zhang[1], Jianxiang Ma[1], Jianqiang Yang[1], Hao Li[1], Feishi Luan[2], Wenfeng Huang[3], Xian Zhang[1,4], Li Yuan[1*†], Ruimin Zhang[5*†] and Chunhua Wei[1*†]

[1]State Key Laboratory of Crop Stress Biology in Arid Areas, College of Horticulture, Northwest A&F University, Yangling, Shaanxi, China, [2]College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, Heilongjiang, China, [3]Vegetable Research Institute of Hainan Academy of Agricultural Sciences, Haikou, Hainan, China, [4]State Key Laboratory of Vegetable Germplasm Innovation, Tianjin, China, [5]College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an, Shandong, China

Male sterility is a valuable trait for watermelon breeding, as watermelon hybrids exhibit obvious heterosis. However, the underlying regulatory mechanism is still largely unknown, especially regarding the related non-coding genes. In the present study, approximately 1035 differentially expressed genes (DEGs), as well as 80 DE-lncRNAs and 10 DE-miRNAs, were identified, with the overwhelming majority down-regulated in male-sterile floral buds. Enrichment analyses revealed that the general phenylpropanoid pathway as well as its related metabolisms was predicted to be altered in a mutant compared to its fertile progenitor. Meanwhile, the conserved genetic pathway *DYT1-TDF1-AMS-MS188-MS1*, as well as the causal gene *ClAMT1* for the male-sterile mutant Se18, was substantially disrupted during male reproductive development. In addition, some targets of the key regulators *AMS* and *MS188* in tapetum development were also down-regulated at a transcriptional level, such as *ABCG26* (*Cla004479*), *ACOS5* (*Cla022956*), *CYP703A2* (*Cla021151*), *PKSA* (*Cla021099*), and *TKPR1* (*Cla002563*). Considering lncRNAs may act as functional endogenous target mimics of miRNAs, competitive endogenous RNA networks were subsequently constructed, with the most complex one containing three DE-miRNAs, two DE-lncRNAs, and 21 DEGs. Collectively, these findings not only contribute to a better understanding of genetic regulatory networks underlying male sterility in watermelon, but also provide valuable candidates for future research.

KEYWORDS

watermelon, male sterility, lncRNAs, miRNAs, regulatory network

Introduction

Male sterility is a common phenomenon in flowering plants that has been widely applied in crop hybrid seed production and heterosis utilization. As a complex biological process, numerous transcription factors have been validated to precisely regulate the development of male reproductive organs (i.e., anthers), such as the tapetum-specific genetic pathway *DYT1-TDF1-AMS-MS188-MS1* (*MS188* is also known as *MYB103*/*MYB80*) (Wilson et al., 2001; Sorensen et al., 2003; Zhang, 2006; Zhang et al., 2007; Zhu et al., 2008; Lu et al., 2020). In this regulatory cascade, *DYT1* (*DYSFUNCTIONAL TAPETUM 1*) and *AMS* (ABORTED MICROSPORES) belong to the bHLH transcription factor (TF) family, *TDF1* (*TAPETAL DEVELOPMENT AND FUNCTION 1*) and *MS188* (*MALE STERILE 188*) encode R2R3 MYB TFs, and *MS1* (*MALE STERILITY 1*) is a TF with leucine zipper-like and PHD-finger motifs. Moreover, proteins encoded by *AMS* and *MS188* functioning as master regulators can activate a series of downstream targets to coordinate pollen development, e.g., *MALE STERILE 2* (*MS2*), *ACYL-COA SYNTHETASE 5* (*ACOS5*), CYTOCHROME P450 genes *CYP703A2* and *CYP704B1*, *POLYKETIDE SYNTHASE A* (*PKSA*) and *PKSB*, *TETRAKETIDE a-PYRONE REDUCTASE1* (*TKPR1*) and *TKPR2* (Xu et al., 2014; Wang et al., 2018; Lu et al., 2020). It is worth mentioning that a similar regulatory pathway has also been verified in tomato and rice pollen development, indicating this is a highly conserved signaling cascade in plants (Jeong et al., 2014; Nan et al., 2017).

Beyond protein-coding genes, non-coding RNAs (ncRNAs) including microRNAs (miRNAs), small interfering RNAs (siRNAs), circular RNAs (circRNAs), and long non-coding RNAs (lncRNAs), have also been confirmed to participate in various aspects of plant growth and development, such as stress-related responses (Hu et al., 2021; Ren et al., 2021; Shi et al., 2021; Wang et al., 2021) and male reproductive development (Li et al., 2015; Yuan et al., 2020). As endogenous non-coding RNAs of 21–25 nt in length, miRNAs mainly contribute to post-transcriptional regulation of downstream targets through mRNA cleavage. For example, *miR156*, *miR159*, and *miR396* can participate in flower development and male fertility in plants, *via* regulating their targets (Millar and Gubler, 2005; Wang et al., 2009; Yuan et al., 2020). With the development of high-throughput sequencing technology, series of differently expressed miRNAs related to male organ development have been identified in various species (Nie et al., 2018; Dhaka et al., 2020). Compared to small miRNAs, lncRNAs are a type of non-coding transcript exceeding 200 nt, and they can be further classified into intergenic lncRNAs (lincRNAs), intronic lncRNAs, and antisense lncRNAs based on their chromosomal locations (Li et al., 2019a; Gao et al., 2020). To date, several lncRNAs have been characterized to function in pollen fertility, such as the lincRNA *BcMF11* in Chinese cabbage (Song et al., 2013) and LDMAR in hybrid rice (Ding et al., 2012). Moreover, approximately 865 differentially expressed lncRNAs were detected to possibly be involved in cotton anther development (Li et al., 2019a). Similarly, potential lncRNAs participating in flower development have also been identified in plants, such as tomato (Yang et al., 2019), strawberry (Kang and Liu, 2015), *Brassica campestris* (Liang et al., 2019), and cotton (Nie et al., 2018; Li et al., 2019a). Interactions among mRNAs, lncRNAs, and miRNAs have revealed that lncRNAs acting as miRNA sponges can competitively bind with miRNAs to indirectly influence the expression of corresponding targets, raising the competitive endogenous RNA (ceRNA) hypothesis (Salmena et al., 2011; Li et al., 2019b; Sun et al., 2020). For example, in watermelon, ceRNA networks of lncRNA/circRNA–miRNA–mRNA interactions, including 23 potential lncRNA–miRNA–mRNA and 125 potential circRNA–miRNA–mRNA interactions, were constructed in response to CGMMV infection (Sun et al., 2020). In maize, ceRNA regulatory networks consisting of 51 known miRNAs, 28 potentially novel miRNAs, 619 ceRNA–miRNA pairs, and 869 miRNA–target gene pairs were proposed to play roles during anther development (Li et al., 2019b). Similarly, correlation networks of lncRNA/circRNA–miRNA–mRNA interactions were inferred to act during flower development in tomato (Yang et al., 2019) and *Brassica campestris* (Liang et al., 2019), respectively.

Watermelon is grown worldwide, and its hybrids exhibit obvious heterosis. Traditionally, the commercial hybrid seeds of watermelon are mainly produced by artificial emasculation and pollination, requiring lots of time and labor. However, the application of male-sterile lines could sufficiently overcome these constraints of hybrid production. To date, several male-sterile mutants have been reported in watermelon (Wei et al., 2021; Zhang et al., 2021), and only one causal gene, *ClATM1*, has been functionally characterized from the male-sterile mutant Se18 previously (Zhang et al., 2021). Compared to the wild type *ClATM1*, its mutant allele contains a 10-bp deletion in the second exon that results in a truncated protein without the bHLH interaction and functional (BIF) domain, leading to the male sterile phenotype in Se18 (Zhang et al., 2021). However, the underlying regulatory network is still poorly understood. In this study, using high-throughput sequencing, we identified differentially expressed genes (DEGs), DE-lncRNAs, and DE-miRNAs at a genome-wide scale from male floral buds between mutant Se18 and its fertile wild-type progenitor. Comparative analyses revealed that the phenylpropanoid-related metabolisms, as well as the highly conserved genetic regulatory pathway *DYT1-TDF1-AMS-MS188-MS1*, were predicted to be disrupted during male reproductive development. Finally, the corresponding ceRNA networks were constructed, providing a foundation for elucidating the complex underlying mechanisms for male sterility in watermelon.

Materials and methods

Plant materials

As described in our previous studies (Wei et al., 2021; Zhang et al., 2021), watermelon mutant Se18 is completely male sterile, and exhibits distinct cytological defects in floral buds at 2.0–2.5 mm in diameter, but without obvious phenotypic differences compared to its fertile wild-type (WT) progenitor. Hence, to precisely investigate

the underlying regulatory network associated with male sterility, floral buds with a length from 2.0 to 2.5 mm were independently sampled from mutant Se18 and WT plants to generate male sterile (Ms) and male fertile (Mf) pools, with three replicates respectively. All the plant materials were grown in the farms of Northwest A&F University, Yangling, China, under natural conditions.

Library preparation and Illumina sequencing

The total RNA were independently extracted from Ms and Mf pools and subsequently sent to Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) for high-throughput sequencing. The chain-specific libraries were constructed for mRNA and lncRNAs, by using the Illumina NovaSeq6000 platform to harvest 150-bp paired-end reads (PE150). The clean data of six libraries (ms_1, ms_2, ms_3 and mf_1, mf_2, mf_3) were mapped onto the watermelon reference genome (97103, V1) using HISAT2 with the parameter '–rna-strandness RF'. Then, the software tools StringTie and Cuffmerge were used to assemble the transcripts.

The small RNA libraries were constructed with the corresponding Ms and Mf pools, and deep sequencing was performed on an Illumina Hiseq^{TM} 2500 platform (Novogene, Beijing, China) to generate 50-bp single-end reads. Clean data were obtained after removing reads containing poly-N runs (N% > 10%), reads with 5′ adapter contaminants, reads without a 3′ adapter or the insert tag, and reads containing poly-A, -T, -G, or -C runs, as well as low-quality reads, from raw data. Then, the clean data from the six libraries (ms_1, ms_2, ms_3 and mf_1, mf_2, mf_3) were mapped onto the watermelon reference genome (97103, V1) using Bowtie (Langmead and Salzberg, 2012), to analyze their expression and chromosome distribution.

Identification of differentially expressed mRNAs, lncRNAs and miRNAs

Based on the fragments per kilobase of transcript per million mapped reads (FPKM), genes with expression changes more than twofold (q-value <0.05) were recognized as differentially expressed genes (DEGs). Using BlastP (e-value = 1e-10), all the DEGs were searched against the *Arabidopsis* protein database (TAIR10) to identify their closest homologs/orthologs. Moreover, transcription factors were also annotated according to the plant transcription factor database PlantTFDB (Jin et al., 2017).

To identify lncRNAs at a genome-wide scale in watermelon, the tools CPC2, CNCI, and PFAM were employed to estimate the coding potential of the assembled transcripts (length > 200 nt, FPKM ≥ 0.5). According to the class codes annotated by the program Cuffcompare, the identified lncRNAs were classified into three types, i.e., intergenic, intronic, and antisense lncRNAs. Similarly, based on the FPKM method, lncRNAs with expression changes more than twofold (q-value <0.05) were recognized as differentially expressed lncRNAs (DE-lncRNAs).

Using miRBase as a reference (specific species *Arabidopsis thaliana*), the mapped small RNA tags were used to identify known miRNAs. For highly accurate identification, only those small RNA tags perfectly matched with known miRNAs were considered as mature sequences of known miRNAs. After removing tags originating from rRNA, tRNA, snRNA, and snoRNA sequences, the remaining small RNA tags were used to identify novel miRNAs with miREvo (Wen et al., 2012) and miRDeep2 (Friedländer et al., 2012). Based on transcripts per million reads (TPM), DESeq was used to identify differentially expressed miRNAs (DE-miRNAs, including known and novel miRNAs) according to the criteria $|\log_2(FC)| \geq 1$ and p-adjust < 0.05. The identified miRNAs with a low total read count number (less than 10) in six libraries were discarded. The chromosome distribution of DEGs, DE-lncRNAs, and DE-miRNAs was visualized using Circos.

Gene ontology and Kyoto encyclopedia of genes and genomes enrichment analysis

To predict the potential functions of lncRNAs, genes located around 100 kb up-stream and down-stream of the DE-lncRNAs were selected as potential *cis*-targets (Ren et al., 2021; Shi et al., 2021). The online tool psRNAtarget and the program psRoot were utilized to predict the potential target genes of DE-miRNAs (Wu et al., 2012; Dai et al., 2018). The subsequent Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DEGs, as well as potential targets of DE-lncRNAs and DE-miRNAs, were performed using OmicShare tools, a free online platform for data analysis (https://www.omicshare.com). Heatmaps were generated using TBtools (Chen et al., 2020).

Construction potential networks among DEGs, DE-lncRNAs, and DE-miRNAs

To infer the potential regulatory relationship between miRNAs and lncRNAs, mature sequences of DE-miRNAs were submitted to the online tool psRNAtarget as templates to predict their target DE-lncRNAs, with default parameters except for a strict Expectation value of 4. The interaction network among DEGs, DE-lncRNAs, and DE-miRNAs was graphically generated by Cytoscape (Shannon et al., 2003).

Validation of DEGs, DE-lncRNAs, and DE-miRNAs by qRT-PCR

To validate the expression of DEGs and DE-lncRNAs, total RNA was extracted from Ms and Mf pools using Trizol® reagent, and the first-strand cDNA was subsequently synthesized *via* the FastKing RT Kit with gDNase (TIANGEN, Beijing, China). Specific primers were designed, and the housekeeping gene *ClACT* (*Cla007792*) was used as an internal reference (Wei et al., 2019; Yue et al., 2021). For DE-miRNAs validation, cDNAs were synthesized *via* the miRcute Plus miRNA First-Strand cDNA Kit,

and qRT-PCR was conducted using the miRcute Plus miRNA qPCR Kit (TIANGEN, Beijing, China). The forward primers were designed based on miRNA-specific sequences, and the universal reverse primer was contained in the qPCR kit. The housekeeping miRNA was U6 snRNA. All the amplification experiments were performed on a StepOnePlus Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). The relative expression level was calculated using the $2^{-\Delta\Delta Ct}$ method as described in our previous studies (Wei et al., 2019; Yue et al., 2021), and significance was determined by Student's *t*-test. All the primers used are listed in Supplementary Supplementary Table 1.

Results

Overview of the whole-transcriptome sequencing dataset

Compared to WT individuals, the male-sterile mutant Se18 showed distinct cytological defects in developmental floral buds with a length from 2.0 to 2.5 mm, such as irregularly shaped and smaller tapetal cells (Wang et al., 2020; Wei et al., 2021). To further uncover the regulatory mechanisms underlying the development of male sterility in watermelon, the corresponding floral buds were independently sampled from Se18 and WT plants and designated as ms_1, ms_2, and ms_3 for male sterility and mf_1, mf_2, and mf_3 for male fertility, respectively, which were subsequently subjected to whole-transcriptome sequencing. The strong correlations among biological replicates suggested that the sequencing data were highly reliable (Supplementary Figure 1A). After filtering the raw data, at least 13.00 Gb of clean data were obtained for each sample (Supplementary Table 2), with Q30 values higher than 91.10% and more than 81.00% of reads successfully mapped to the reference genome 97103 (V1). Following the filter criteria (Supplementary Figure 2A), 36,873 lncRNAs were finally identified (Supplementary Figure 2B), with the majority being lincRNAs (Supplementary Figure 2C). Compared to protein-coding transcripts (mRNAs), most lncRNAs were characterized to have no more than two exons and shorter sequence lengths (Supplementary Figures 2D, E).

To detect the responsive sRNAs involved in male sterility in watermelon, six libraries were correspondingly constructed for sequencing, obtaining approximately 300 Mb of clean data for each sample with a mapping ratio higher than 88.0% (Supplementary Table 3). The reliability of sequencing data among biological replicates was assessed using the Pearson correlation coefficient (Supplementary Figure 1B). Further analysis revealed that the size of clean reads mapped onto the genome mainly ranged from 21 to 25 nt, with 24 nt representing the largest class occupying more than 50% in each library (Figure 1A), which is similar to the proportions observed in other plants, such as *Arabidopsis* and cotton (Fahlgren et al., 2010; Xing et al., 2014; Yu et al., 2020). Using the database miRBase20.0, a set of 26 known miRNA families were identified and contained approximately 46 members, with the largest two families miRNA396 and miRNA399 harboring four isoforms respectively (Figure 1B, Supplementary Table 4). Notably, 39 out of the 46 known miRNAs were 21-nt miRNAs, and the majority (69.23%) contained 'U' at the first position (Figure 1C, Supplementary Table 4). Additionally, 67 novel miRNAs were also identified using miREvo and mirDeep2 with default parameters (Figure 1D, Supplementary Table 5), with the 24-nt miRNAs as the most abundant class (53.73%).

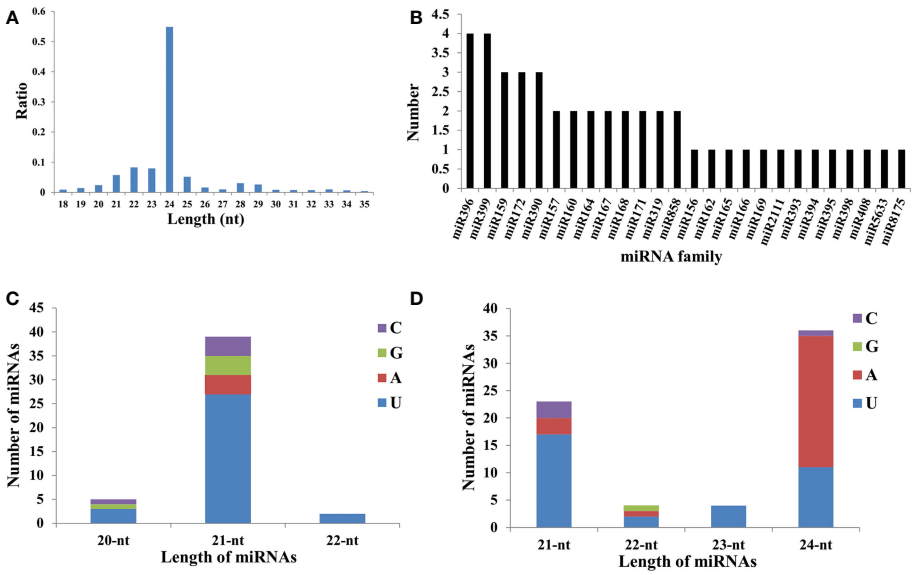


FIGURE 1
Detailed information for the identified miRNAs. **(A)** The length distribution of small RNA read sequences. **(B)** The member statistics in known miRNA families. The length distribution of known miRNAs **(C)** and novel miRNAs **(D)**. The four nucleotides (C, G, A, and U) are represented in purple, green, red, and blue respectively.

Differential expression analyses of mRNAs, lncRNAs, and miRNAs

Analysis of DEGs

Among all the detected mRNAs, a total of 1035 DEGs were identified, containing 350 up-regulated and 685 down-regulated genes in the male-sterile mutant Se18 (Figure 2). Referring to the proteomic data published previously (Wei et al., 2021), about 164 DEGs were also detected with significant changes at the protein level (Supplementary Table 6). Moreover, there were three DEGs (*Cla013360*, *Cla013805*, and *Cla021282*) with increased abundance at the transcription level, but reduced protein accumulation in Se18 (Supplementary Figure 3A). To further predict the potential function, all 1035 DEGs were searched against the *Arabidopsis* protein database using BlastP. Thus, 961 DEGs were successfully matched with their best homologies (Supplementary Table 6), including the functionally characterized tapetum-specific genes *DYT1* (*Cla010083*), *AMS* (*Cla015818*), *TDF1* (*Cla019144*), and *bHLH091* (*Cla010576*).

The GO functional categorization analysis revealed that approximately 266 DEGs, including 68 up-regulated and 198 down-regulated genes, were significantly enriched into 39 GO terms across the three categories biological process (BP, 22), cellular component (CC, 2), and molecular function (MF, 15) (Supplementary Table 7). Notably, 18 out of the 22 BP GO terms ultimately corresponded to four terms, 'stilbene biosynthetic process' (GO:0009811), 'lignin biosynthetic process' (GO:0009809), 'coumarin biosynthetic process' (GO:0009805), and 'pollen wall assembly' (GO:0010208) (Supplementary Figures 3B, 4). Moreover, KEGG enrichment analysis showed that 55 up-regulated and 112 down-regulated DEGs were classified into 19 different pathways at a *P*-value ≤ 0.05 threshold (Supplementary Table 8). The top five significantly enriched metabolic pathways were as follows: Biosynthesis of secondary metabolites (ko01110; 79, 47.30%), Flavonoid biosynthesis (ko00941; 8, 4.79%), Phenylpropanoid biosynthesis (ko00940; 21, 12.57%), Metabolic pathways (ko01100; 113, 67.66%), and Galactose metabolism (ko00052; 9, 5.39%). Taken together, these enrichment analyses suggested that the phenylpropanoid pathway, as well as the related metabolisms, was predicted to be altered in mutant Se18, which is consistent with our previously published results (Wei et al., 2021).

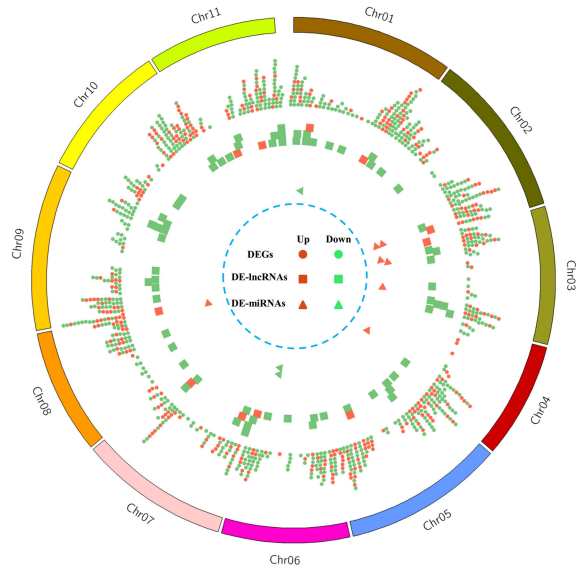


FIGURE 2
The chromosomal distribution of DEGs, DE-lncRNAs, and DE-miRNAs. From the outside to inside, the circles are watermelon chromosomes, distribution of the up-regulated (red) and down-regulated (green) DEGs (dot), DE-lncRNAs (square), and DE-miRNAs (triangle).

Analysis of DE-lncRNAs

To uncover the lncRNAs potentially involved in watermelon male sterility, using parameters $|\log_2(FC)| \geq 1$ and q-value < 0.05 as thresholds, a total of 80 DE-lncRNAs were identified (Figure 2, Supplementary Table 9), including 11 up-regulated and 69 down-regulated lncRNAs in mutant Se18. Following published studies (Ren et al., 2021; Shi et al., 2021), we characterized the genes that are *cis*-targets of the DE-lncRNAs, which are located around 100 kb up-stream and down-stream of the DE-lncRNAs. After removing redundant targets, a set of 1247 genes was obtained, comprising 1345 mRNA–DE-lncRNA pairs (Supplementary Table 10). To further elucidate the potential functions of DE-lncRNAs, GO enrichment analysis was performed with the 1247 *cis*-target genes, revealing that only 50 genes were significantly categorized into 8 MF and 11 BP GO terms (Supplementary Table 7). Notably, the terms 'phenylalanine ammonia-lyase activity' (GO:0045548) in MF and 'L-phenylalanine catabolic process (GO:0006559) in BP were enriched, suggesting the potential functions of the related DE-lncRNAs in the phenylalanine metabolic process. Additionally, KEGG pathway enrichment analysis showed that only 33 targets were enriched in six different pathways (Supplementary Table 8), including the phenylalanine metabolism (ko00360; 9, 27.27%), in line with the GO analysis.

Combined with the RNA-seq data, approximately 78 out of the 1247 *cis*-target genes were identified as DEGs (Supplementary Table 10), including eight transcription factors such as *Cla002749* (bHLH), *Cla009235* (WRKY), and *Cla021810* (MYB). GO functional categorization analysis revealed that only 7 up-regulated and 13 down-regulated *cis*-targets were significantly enriched in approximately 32 GO terms (Supplementary Table 7). Notably, the MF terms were mainly related to protein methyltransferase activity, such as histone-lysine *N*-methyltransferase activity (GO:0018024) and protein-lysine *N*-methyltransferase activity (GO:0016279), while BP terms were involved in the lysine catabolic process (e.g., terms GO:0006554, GO:0006553, GO:0009068), terpenoid biosynthetic process (e.g., terms GO:0016104, GO:0006722, GO:0019742), and lipid metabolic process (e.g., terms GO:0006629, GO:0044255) (Supplementary Figures 3C, D). Consistently, KEGG enrichment analysis showed that terpenoid backbone biosynthesis (ko00900) and glycerolipid metabolism (ko00561) were also significantly enriched (Supplementary Table 8), inferring that lncRNAs might regulate genes involved in synthesis and modification of amino acids, as well as the lipid metabolic process.

Analysis of DE-miRNAs

To identify the potential miRNAs involved in male sterility, expression of all the detected miRNAs was estimated by the TPM approach and subsequently compared between WT and mutant Se18 sequence data. As a result, only 10 DE-miRNAs were obtained (Figure 2, Supplementary Table 11), including four known and six novel miRNAs. Among the known miRNAs, both of the miR396 members were down-regulated in the male-sterile mutant and predicted to regulate more than 200 targets (Supplementary Table 11). However, only about ten targets were identified as DEGs according to the RNA-seq data. In contrast, five out of six novel DE-miRNAs were up-regulated in Se18, with dozens of targets but only few characterized as DEGs. Additionally, 18 DE-lncRNAs were predicted as potential targets of 30 miRNAs, but only three were recognized as DE-miRNAs (Supplementary Table 12).

Validation of DE-mRNAs, DE-miRNAs, and DE-lncRNAs by qRT-PCR

To validate the differential expression of mRNAs, lncRNAs, and miRNAs based on the multi-omics analysis, qRT-PCR was utilized to confirm the expression patterns between male fertile and sterile buds. As a result, the transcriptional levels of ten selected DEGs were consistent with the bioinformatic analysis (Supplementary Table 13), which have also been confirmed to match with their protein accumulations in our previous published research (Wei et al., 2021). Moreover, expression patterns of all ten DE-miRNAs and ten randomly selected DE-lncRNAs were generally concordant with the bioinformatic results (Figure 3), suggesting the reliability of the high-throughput sequencing data.

Construction of miRNA–lncRNA–mRNA regulatory networks

As mentioned above, there were 82 potential DElncRNA–DEG (including 53 DE-lncRNAs and 78 DEGs, Supplementary Table 10), 30 potential DEmiRNA–DEG (including 8 DE-miRNAs and 23 DEGs, Supplementary Table 11), and 3 potential DEmiRNA–DElncRNA (including 3 DE-miRNAs and 2 DE-lncRNAs, Supplementary Table 12) interactions. Among them, the majority were potential one-to-one interactions, while the rest were potential one-to-many interactions, such as LNC_014758, which is predicted to regulate three up-regulated genes (*Cla013358*, *Cla013360*, *Cla013371*) and one down-regulated gene *Cla013363* (Supplementary Figure 5). Moreover, we only identified two potential miRNA–lncRNA–mRNA interactions (Figure 4). Compared to the Cl-miR171a–LNC_004095–Cla011387 interaction, the key regulatory network was much more complex, containing three DE-miRNAs (up-regulated nove_71 and down-regulated Cl-miR396a-5p and Cl-miR396b-5p), two DE-lncRNAs (down-regulated LNC_011135 and LNC_035385), and 21 DEGs including MYB transcription factor *Cla007663* and GROWTH REGULATING FACTOR (GRF) *Cla016859*. Among these DEGs, one up-regulated gene (*Cla014800*) and four down-regulated genes (*Cla010475*, *Cla012139*, *Cla013727*, and *Cla016859*) were predicted to be regulated by both Cl-miR396a-5p and Cl-miR396b-5p (Figure 4). Additionally, the down-regulated LNC_011135 was also detected as a potential target of two miRNAs, which was predicted to be regulated by *Cla016682* expression; however, the other down-regulated LNC_035385 was predicted to regulate three DEGs, including the genes *Cla016358* and *Cla016362* with decreased transcriptional abundance as well as a presumable target (*Cla016373*) of Cl-miR396b-5p with an increased transcript level.

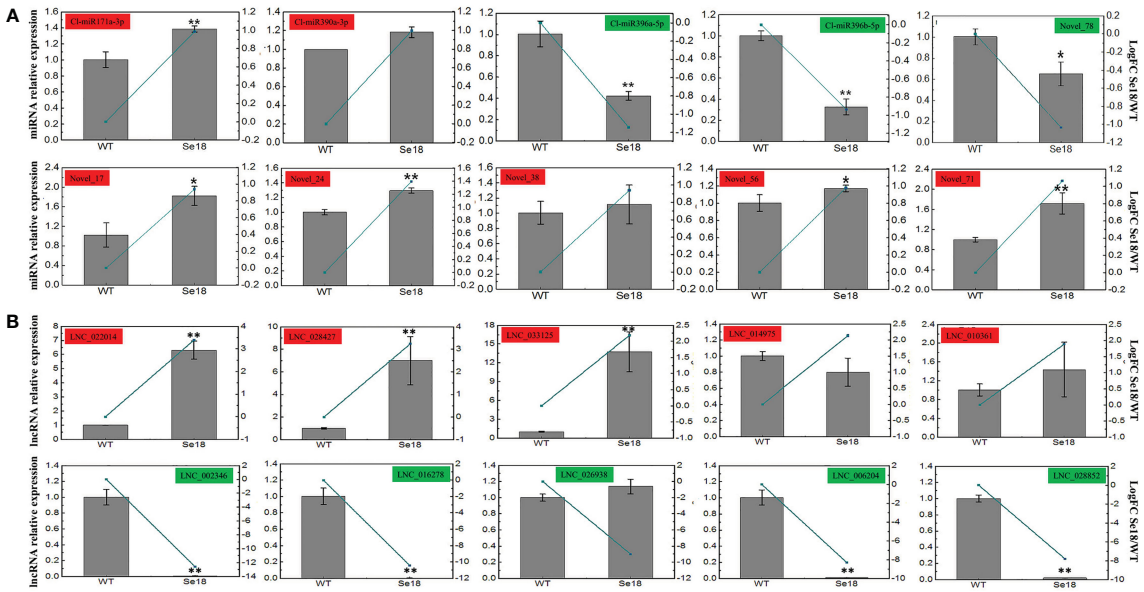


FIGURE 3
Expression validation of DE-miRNAs **(A)** and DE-lncRNAs **(B)** by qRT-PCR. Gray bars and green lines represent the relative expression and bioinformatics results respectively. The relative expression levels of each gene in WT are normalized as reference and presented as means ± SD (three biological replicates, *$P < 0.05$, **$P < 0.01$, Student's *t*-test).

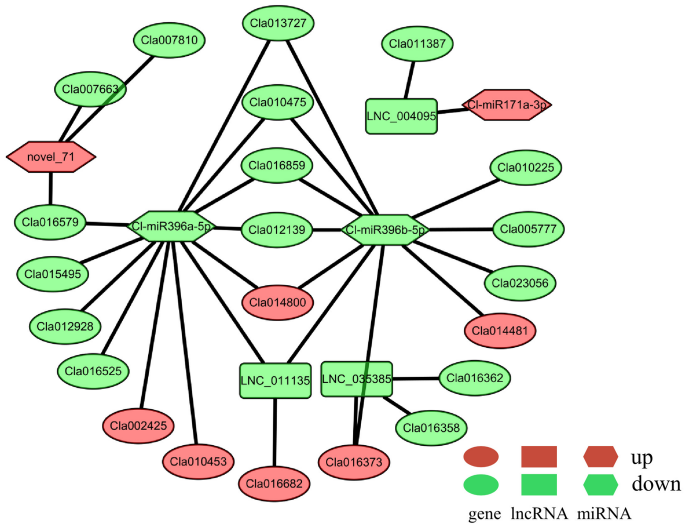


FIGURE 4
ceRNA regulatory networks predictably involved in male sterility of watermelon. DEGs, DE-lncRNAs, and DE-miRNAs are indicated as ellipses, rectangular, and hexagon respectively. The red color represents the up-regulated expression, while the green represents the down-regulated expression.

Identification of transcription factors possibly involved in male sterility

Based on the plant transcription factor database PlantTFDB, approximately 54 up-regulated and 45 down-regulated DEGs were annotated as transcription factors, belonging to 25 families (Figure 5A). The MYB TF family, the largest one among them, contained 18 members, followed by bHLH (10), ERF (9), and NAC (8). Additionally, the majority of MYB TFs (15 out of 18) were enriched in the KEGG pathway 'Plant-pathogen interaction' (ko04626) (Supplementary Table 8). Combined with the analyses of DE-lncRNAs and DE-miRNAs, eight TFs were predicted to be regulated by DE-lncRNAs, and two TFs were possibly targeted by DE-miRNAs (Figure 5B). Notably, the bHLH TF *Cla002749* and GRF TF *Cla016859* were predicted to be regulated by two DE-lncRNAs and DE-miRNAs, respectively (Supplementary Table 14).

DEGs involved in the phenylpropanoid biosynthesis and its related metabolism pathways

As revealed by previous studies (Vogt, 2010; Yadav et al., 2020; Wei et al., 2021), the phenylpropanoid biosynthesis pathway, serving as a major specialized metabolism in plants, can provide various components for pollen and anther development, including

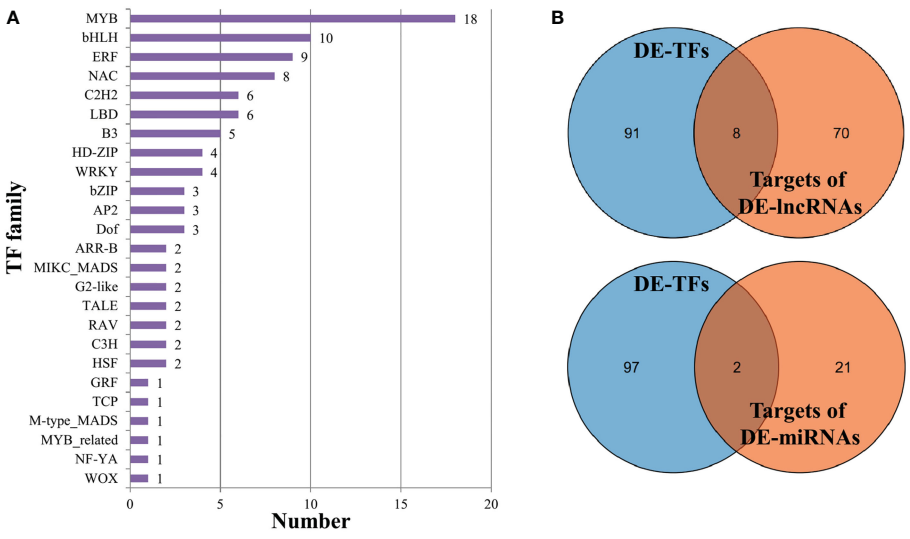


FIGURE 5
Transcription factors annotated in DEGs. **(A)** The members in each TF family. **(B)** The intersection of DE-TFs and targets of DE-lncRNAs (or DE-miRNAs).

coumarins, flavonoids, and lignins. According to the KEGG pathway analysis (Supplementary Table 8), phenylpropanoid biosynthesis (ko00940), along with phenylalanine metabolism (ko00360) and flavonoid biosynthesis (ko00941), was significantly enriched among the 1035 DEGs. For example, the *PAL* gene (*Cla018300*) encoding the enzyme phenylalanine ammonia lyase, with the function of catalyzing the amino acid phenylalanine into *trans*-cinnamic acid (Figure 6), was significantly down-regulated in male-sterile floral buds (log_2(FC) = -1.01, Supplementary Table 6) and was also detected with considerably decreased abundance at protein level (Wei et al., 2021). Moreover, the accumulations of other key genes in the general phenylpropanoid biosynthesis pathway, e.g., *C4H* (*Cla005785* and *Cla005787*), *4CL2* (*Cla015295*, *Cla015298*, *Cla017226*, and *Cla018820*), *ACOS5* (*Cla022956*), *HCT* (*Cla022714*), *CCoAOMT* (*Cla016012*), were significantly altered in the male-sterile mutant Se18 (Figure 6). The intermediate CoA esters from the general phenylpropanoid biosynthesis pathway are subsequently synthesized into various lignins, *via* the actions of the enzymes CCR1 (Cla017209), MEE23 (Cla009844), OTM1 (Cla004454), PRX40 (Cla009234), PRX53 (Cla003191), and PER64 (Cla022405). As important intermediate constituents, CoA esters can be further transmitted into the flavonoid biosynthesis pathway (Falcone Ferreyra et al., 2012; Fellenberg and Vogt, 2015; Shi et al., 2015), in which cinnamoyl CoA and p-coumaroyl-CoA are catalyzed into pinocembrin chalcone and naringenin chalcone by chalcone synthase CHS (*Cla002945* with a -11.50-fold mRNA level in mutant Se18), respectively. As expected, the genes encoding enzymes that convert chalcones into the flavonols galangin and kaempferol were significantly down-regulated, including *CHIL* (*Cla007294*, -1.10-fold), *F3H* (*Cla008896*, -5.43-fold), and *FLS* (*Cla008008*, -2.81-fold) (Figure 6, Supplementary Table 6). Collectively, the conserved general phenylpropanoid biosynthesis pathway, as well as its related metabolism pathways, was inferred to undergo considerable disruption during another development in the male-sterile mutant Se18.

DEGs involved in regulatory cascade in the nucleus

The transcriptional cascade *DYT1–TDF1–AMS–MS188–MS1* is considered to precisely regulate tapetum development and mature pollen production in plants (Wilson et al., 2001; Sorensen et al., 2003; Zhang, 2006; Zhang et al., 2007; Zhu et al., 2008; Jeong et al., 2014; Lu et al., 2020). Previously, three other bHLH TFs (bHLH010, bHLH089, bHLH091) were thoroughly characterized to function in pollen fertility, through interacting with DYT1 and/or AMS to activate downstream targets (Cui et al., 2016; Ferguson et al., 2017). In the watermelon male-sterile mutant Se18, the causal gene *ClATM1* (*Cla010576*) was recently cloned and identified as the ortholog of *Arabidopsis bHLH010*, *bHLH089*, and *bHLH091*, which is down-regulated in male-sterile flower of Se18 (Zhang et al., 2021), consistent with our transcriptome data (Figure 7). Meanwhile, all the other TFs in the regulatory cascade were decreased at a transcriptional level in mutant Se18, except for the gene *DYT1* (*Cla010083*). Functioning as the key regulators in this pathway (Xu

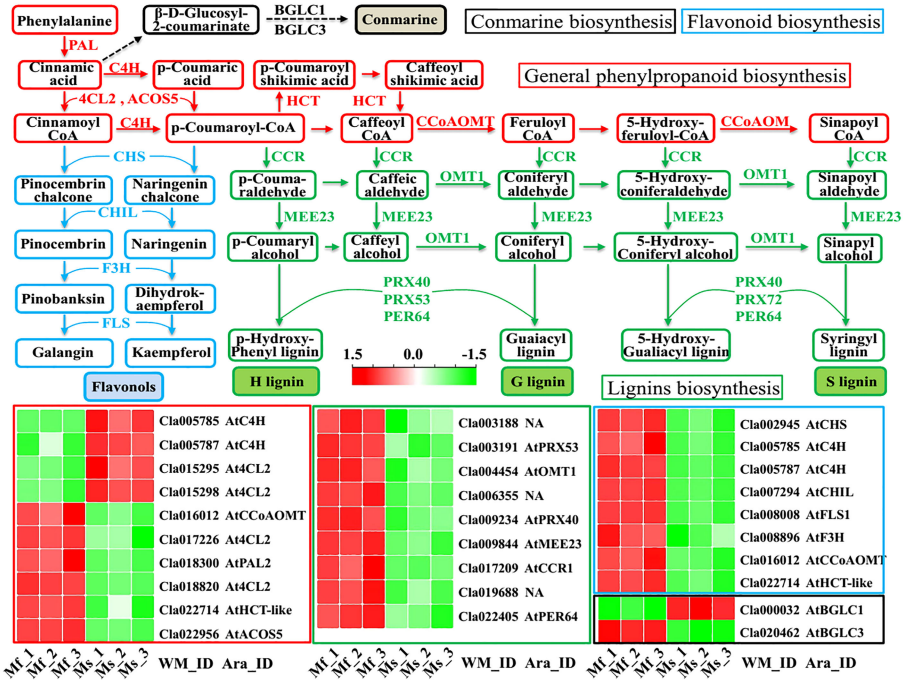


FIGURE 6
The general phenylpropanoid biosynthesis and its related secondary metabolisms enriched by KEGG analysis. Lines in different colors are indexed as followed: Red, general phenylpropanoid biosynthesis; Blue, chalcone and flavonol biosynthesis; Green, lignins biosynthesis; Black, conmarine biosynthesis.

et al., 2014; Wang et al., 2018; Lu et al., 2020), some targets of *AMS* and *MS188* were also down-regulated at the transcriptional level, such as *QRT2* (*Cla009870*), *ABCG26* (*Cla004479*), *ACOS5* (*Cla022956*), *CYP703A2* (*Cla021151*), *PKSA* (*Cla021099*), and *TKPR1* (*Cla002563*), demonstrating the obvious disruption of this conserved regulatory cascade in Se18.

Discussion

Male flower development is a complex biological process regulated by precise networks, and dysfunction of any controlling genes possibly leads to male sterility. For example, a 10-bp sequence deletion in the bHLH gene *ClATM1* resulted in a complete male-sterile phenotype of the watermelon mutant Se18 (Zhang et al., 2021). Furthermore, according to our transcriptomic and TMT-based proteomic analyses, lipid metabolism, as well as the phenylpropanoid pathway, is predicted to be altered in mutant Se18 compared to its fertile progenitor (Wang et al., 2020; Wei et al., 2021). In the present study, the general phenylpropanoid pathway, together with its related metabolisms, was also impaired in male-sterile floral buds, as expected (Figure 6, Supplementary Table 8). Importantly, 14 out of the 25 DEGs enriched in phenylpropanoid-related pathways were also detected to have significant abundance changes at the protein level (Supplementary Table 6), further confirming the disruption of the corresponding metabolisms in male-sterile flower development. Given that the general phenylpropanoid pathway, together with its related secondary metabolisms, produces numerous substances, serving as important components for pollen wall development (e.g., lignin, coumarins, stilbenes, and flavonoids) (Vogt, 2010; Fellenberg and Vogt, 2015; Shi et al., 2015; Yadav et al., 2020; Wei et al., 2021), it is not surprising to detect substantial differences in plants with male reproductive defects, such as autotetraploid watermelon (Yi et al., 2022), tomato (Yang et al., 2019), *Brassica campestris* (Liang et al., 2019; Tang et al., 2022), cotton (Li et al., 2019a), and alfalfa (Xu et al., 2022).

In plants, non-coding RNAs (ncRNAs), including lncRNAs and miRNAs, have been determined to play essential roles in male reproductive development (Li et al., 2015). For example, the 828-bp lncRNA *BcMF11* from Chinese cabbage has been reported to function in pollen fertility (Song et al., 2013). Similarly, in hybrid rice, the sufficient abundance of the lincRNA LDMAR (1236 nt in length) is required for normal pollen formation under long-day conditions (Ding et al., 2012). In contrast with lncRNAs, miRNAs are endogenous single-stranded ncRNAs of 21 to 25 nt in length that function as post-transcriptional regulators in another development. For example, the conserved miR156 can directly target SPL genes to maintain anther fertility (Wang et al., 2009), while miR159 negatively regulates MYB members (e.g., *MYB33* and *MYB65*) to affect male sterility (Millar and Gubler, 2005). Recently, using high-throughput sequencing technologies, numerous candidate lncRNAs, as well as miRNAs, have been identified with potential roles in flower development, such as in tomato (Yang et al., 2019), *Brassica campestris* (Liang et al., 2019), cotton (Nie et al., 2018; Li et al., 2019a), and maize (Li et al., 2019b). In the present study, we identified a total of 80 DE-lncRNAs and 10 DE-miRNAs, which were predominantly down-regulated in the male-sterile mutant Se18 (Figure 2, Supplementary Tables 9, 11). Consistently, the lipid- and phenylalanine-related metabolisms were also enriched, according to the KEGG analysis of genes potentially targeted by DE-lncRNAs (Supplementary Table 8).

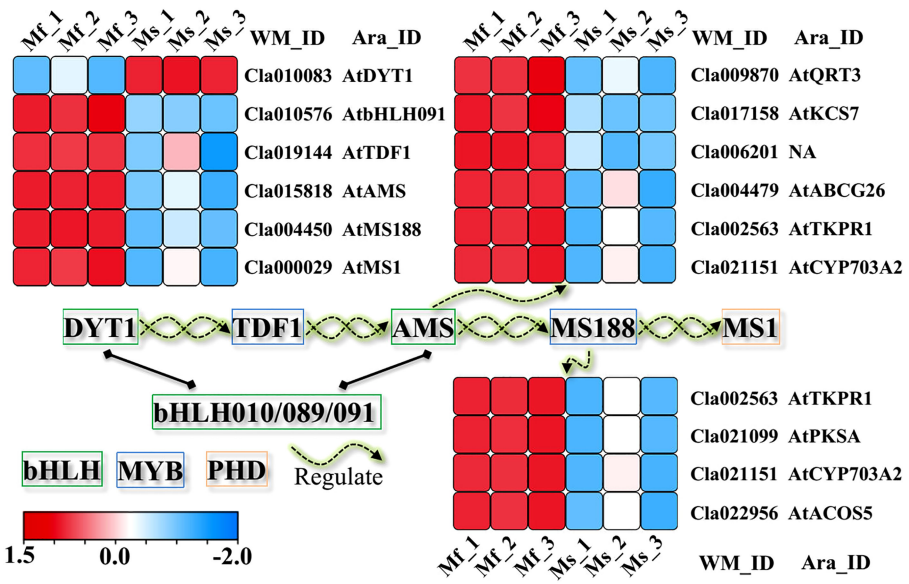


FIGURE 7
The disrupted genetic regulatory cascade in controlling tapetum development. All the TFs as well as the downstream targets of *AMS* (Xu et al., 2014) and *MS188* (Wang et al., 2018), were down-regulated in floral buds of mutant Se18, except for gene *DYT1* (*Cla010083*).

The interaction between evolutionarily conserved miR396 and its target GRF genes has been confirmed to be involved in plant development during vegetative and reproductive stages (Li et al., 2015; Yuan et al., 2020). Additionally, overexpression of miR396 was sufficient to induce pollen sterility (Yuan et al., 2020). However, in the watermelon mutant Se18, two miR396 members (Cl-miR396a and Cl-miR396b) were both detected to have decreased accumulations in male-sterile floral buds (Supplementary Table 11). Additionally, the two Cl-miR396 members, together with other DE-miRNAs, DE-lncRNAs, and DEGs, were predicted to form a complex miRNA–lncRNA–mRNAs regulatory regulatory network (Figure 4), to participate in male reproductive development in watermelon. Similar to the complex ceRNA networks in watermelon after cucumber green mottle mosaic virus (CGMMV) infection (Sun et al., 2020), diverse expression trends were also observed in this research (Figure 4, Supplementary Figure 5). For example, both Cl-miR396a and Cl-miR396b were down-regulated in male-sterile floral buds, whereas some of their predicted targets (e.g., *Cla014800* and *Cla016682*) were up-regulated, though others were down-regulated, including *Cla012139* and *Cla016859*. These results suggested the complexity of the ceRNA networks involved in male reproductive development in watermelon.

In plants, several transcription factors have been reported to be associated with tapetal function and pollen development, including two bHLH genes (*DYT1* and *AMS*), two MYB genes (*TDF1* and *MS188*) and one PHD gene (*MS1*) (Jeong et al., 2014; Nan et al., 2017). In addition, the proteins bHLH010, bHLH089, and bHLH091 have been confirmed to interact with DYT1 and AMS, respectively (Xu et al., 2010; Cui et al., 2016), and further activate *DYT1* through feed-forward and positive feedback loops (Cui et al., 2016). Based on the present transcriptome analyses, only *DYT1* (*Cla010083*) was detected with up-regulated expression in the watermelon mutant Se18, while the other four TFs were down-regulated (Figure 7, Supplementary Table 6). In addition, *AMS* and *MS188* function as key regulators (Xu et al., 2014; Wang et al., 2018; Lu et al., 2020), some targets of which were also expected to be decreased in mutant male-sterile flower buds, including *QRT2* (*Cla009870*), *ABCG26* (*Cla004479*), *ACOS5* (*Cla022956*), *CYP703A2* (*Cla021151*), *PKSA* (*Cla021099*), and *TKPR1* (*Cla002563*). Furthermore, approximately 99 differentially expressed TFs were identified, belonging to 25 families (Supplementary Table 14). Notably, the bHLH transcription factor *Cla002749* and GRF TF *Cla016859* were predicted to be regulated by two DE-lncRNAs and DE-miRNAs respectively, providing valuable candidates for further research. Previously, *ClAMT1*, the causal gene for male sterility in Se18, was functionally characterized (Zhang et al., 2021), revealing it displays a clear dose effect and down-regulation in the Se18 mutant, consistent with the transcriptome data (Figure 7). However, we did not detect any miRNAs or lncRNAs that regulate *ClAMT1* according to our bioinformatic analyses. Thus, this study offers new insights into the regulatory mechanisms underlying male sterility in watermelon and provides valuable candidates for further exploration of the complex regulatory networks.

Data availability statement

The data presented in the study are deposited in the NCBI database repository, accession number PRJNA926906.

Author contributions

YZ, PXN: Conceptualization. YL, ZRM, WCH: Methodology. LJY, SFF, YLJ: Software. HYJ, CXY, LX: Validation. YZ, PXN: Formal analysis. ZY, MJX: Investigation. YJQ, LFS: Resources. YZ: Roles/Writing—original draft. LH, ZRM, WCH: Writing & editing. WCH: Visualization. YL, ZRM, WCH: Supervision. HWF, ZX, WCH: Funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Research and Development Project of Yangling Seed Industry Innovation Center (Ylzy-sc-01), the Modern Agro-Industry Technology Research System of China (No.CARS-25), and the Innovation Center for Academician Teams of Hainan Province.

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.2023.1138415/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Pearson correlation analyses of transcriptome **(A)** and small RNA **(B)** sequencing data among biological replicates. The color scale indicates the strength of the correlation, and the deeper color reflects higher correlation.

SUPPLEMENTARY FIGURE 2
Detail information of lncRNAs. **(A)** The basic identification steps for lncRNAs. **(B)** Screening non-coding lncRNAs using three databases (CPC2, CNCI, and

PFAM). **(C)** Classification of lncRNAs. Comparison of exon number **(D)** and length **(E)** between mRNAs and lncRNAs.

SUPPLEMENTARY FIGURE 3

Comparative and GO enrichment analyses of DEGs. **(A)** The abundance of DEGs at transcriptional and protein levels. Brief schematic diagram of GO terms of DEGs in BP category **(B),** as well as differentially expressed *cis*-targets of DE-lincRNAs in MF **(C)** and BP **(D)** categories. GO ID, term description, *p*-adjust value, genes in input and background lists were shown in each term.

SUPPLEMENTARY FIGURE 4

Detailed GO terms of DEGs in BP category. The detailed information of GO terms was displayed in following order: GO term, term description, *p*-adjust value, genes in input and background lists. The colors represented the level of significance in each term.

SUPPLEMENTARY FIGURE 5

Potential regulatory networks of DEGs, DE-lncRNAs, and DE-miRNAs. DEGs, DE-lncRNAs, and DE-miRNAs are indicated as ellipses, rectangular, and hexagon respectively. The red color represents the up-regulated expression, while the green represents the down-regulated expression.

References

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

Cui, J., You, C., Zhu, E., Huang, Q., Ma, H., and Chang, F. (2016). Feedback regulation of *DYT1* by interactions with downstream bHLH factors promotes *DYT1* nuclear localization and anther development. *Plant Cell* 28 (5), 1078–1093. doi: 10.1105/tpc.15.00986

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server, (2017 release). *Nucleic Acids Res.* 46 (W1), W49–W54. doi: 10.1093/nar/gky316

Dhaka, N., Sharma, S., Vashisht, I., Kandpal, M., Sharma, M. K., and Sharma, R. (2020). Small RNA profiling from meiotic and post-meiotic anthers reveals prospective miRNA-target modules for engineering male fertility in sorghum. *Genomics* 112 (2), 1598–1610. doi: 10.1016/j.ygeno.2019.09.009

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U.S.A.* 109 (7), 2654–2659. doi: 10.1073/pnas.1121374109

Fahlgren, N., Jogdeo, S., Kasschau, K. D., Sullivan, C. M., Chapman, E. J., Laubinger, S., et al. (2010). MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22 (4), 1074–1089. doi: 10.1105/tpc.110.073999

Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00222

Fellenberg, C., and Vogt, T. (2015). Evolutionarily conserved phenylpropanoid pattern on angiosperm pollen. *Trends Plant Sci.* 20 (4), 212–218. doi: 10.1016/j.tplants.2015.01.011

Ferguson, A. C., Pearce, S., Band, L. R., Yang, C., Ferjentsikova, I., King, J., et al. (2017). Biphasic regulation of the transcription factor *ABORTED MICROSPORES* (*AMS*) is essential for tapetum and pollen development in arabidopsis. *New Phytol.* 213 (2), 778–790. doi: 10.1111/nph.14200

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40 (1), 37–52. doi: 10.1093/nar/gkr688

Gao, C., Sun, J., Dong, Y., Wang, C., Xiao, S., Mo, L., et al. (2020). Comparative transcriptome analysis uncovers regulatory roles of long non-coding RNAs involved in resistance to powdery mildew in melon. *BMC Genomics* 21 (1), 125. doi: 10.1186/s12864-020-6546-8

Hu, J., Lan, M., Xu, X., Yang, H., Zhang, L., Lv, F., et al. (2021). Transcriptome profiling reveals molecular changes during flower development between male sterile and fertile Chinese cabbage (*Brassica rapa* ssp. pekinensis) lines. *Life (Basel)* 11 (6), 525. doi: 10.3390/life11060525

Jeong, H. J., Kang, J. H., Zhao, M., Kwon, J. K., Choi, H. S., Bae, J. H., et al. (2014). Tomato *male sterile 10*^{35} is essential for pollen development and meiosis in anthers. *J. Exp. Bot.* 65 (22), 6693–6709. doi: 10.1093/jxb/eru389

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45 (D1), D1040–D1045. doi: 10.1093/nar/gkw982

Kang, C., and Liu, Z. (2015). Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics* 16, 815. doi: 10.1186/s12864-015-2014-2

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9 (4), 357–359. doi: 10.1038/nmeth.1923

Li, Y., Qin, T., Dong, N., Wei, C., Zhang, Y., Sun, R., et al. (2019a). Integrative analysis of the lncRNA and mRNA transcriptome revealed genes and pathways potentially involved in the anther abortion of cotton (*Gossypium hirsutum* l.). *Genes (Basel)* 10 (12), 947. doi: 10.3390/genes10120947

Li, Z., An, X., Zhu, T., Yan, T., Wu, S., Tian, Y., et al. (2019b). Discovering and constructing ceRNA-miRNA-target gene regulatory networks during anther development in maize. *Int. J. Mol. Sci.* 20 (14), 3480. doi: 10.3390/ijms20143480

Li, Z. F., Zhang, Y. C., and Chen, Y. Q. (2015). miRNAs and lncRNAs in reproductive development. *Plant Sci.* 238, 46–52. doi: 10.1016/j.plantsci.2015.05.017

Liang, Y., Zhang, Y., Xu, L., Zhou, D., Jin, Z., Zhou, H., et al. (2019). CircRNA expression pattern and ceRNA and miRNA-mRNA networks involved in anther development in the CMS line of *Brassica campestris*. int. *J. Mol. Sci.* 20 (19), 4808. doi: 10.3390/ijms20194808

Lu, J. Y., Xiong, S. X., Yin, W., Teng, X. D., Lou, Y., Zhu, J., et al. (2020). *MS1*, a direct target of *MS188*, regulates the expression of key sporophytic pollen coat protein genes in arabidopsis. *J. Exp. Bot.* 71 (16), 4877–4889. doi: 10.1093/jxb/eraa219

Millar, A. A., and Gubler, F. (2005). The arabidopsis GAMYB-like genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17 (3), 705–721. doi: 10.1105/tpc.104.027920

Nan, G. L., Zhai, J., Arikit, S., Morrow, D., Fernandes, J., Mai, L., et al. (2017). *MS23*, a master basic helix-loop-helix factor, regulates the specification and development of the tapetum in maize. *Development* 144 (1), 163–172. doi: 10.1242/dev.140673

Nie, H., Wang, Y., Su, Y., and Hua, J. (2018). Exploration of miRNAs and target genes of cytoplasmic male sterility line in cotton during flower bud development. *Funct. Integr. Genomics* 18 (4), 457–476. doi: 10.1007/s10142-018-0606-z

Ren, J., Jiang, C., Zhang, H., Shi, X., Ai, X., Li, R., et al. (2021). LncRNA-mediated ceRNA networks provide novel potential biomarkers for peanut drought tolerance. *Physiol. Plant*, 174(1):e13610. doi: 10.1111/ppl.13610

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* 146 (3), 353–358. doi: 10.1016/j.cell.2011.07.014

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shi, F., Xu, H., Liu, C., Tan, C., Ren, J., Ye, X., et al. (2021). Whole-transcriptome sequencing reveals a vernalization-related ceRNA regulatory network in chinese cabbage (*Brassica campestris* l. ssp. pekinensis). *BMC Genomics* 22 (1), 819. doi: 10.1186/s12864-021-08110-2

Shi, J., Cui, M., Yang, L., Kim, Y. J., and Zhang, D. (2015). Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci.* 20 (11), 741–753. doi: 10.1016/j.tplants.2015.07.010

Song, J. H., Cao, J. S., and Wang, C. G. (2013). *BcMF11*, a novel non-coding RNA gene from brassica campestris, is required for pollen development and male fertility. *Plant Cell Rep.* 32 (1), 21–30. doi: 10.1007/s00299-012-1337-6

Sorensen, A. M., Kröber, S., Unte, U. S., Huijser, P., Dekker, K., and Saedler, H. (2003). The arabidopsis *ABORTED MICROSPORES* (*AMS*) gene encodes a MYC class transcription factor. *Plant J.* 33 (2), 413–423. doi: 10.1046/j.1365-313x.2003.01644.x

Sun, Y., Zhang, H., Fan, M., He, Y., and Guo, P. (2020). Genome-wide identification of long non-coding RNAs and circular RNAs reveal their ceRNA networks in response to cucumber green mottle mosaic virus infection in watermelon. *Arch. Virol.* 165 (5), 1177–1190. doi: 10.1007/s00705-020-04589-4

Tang, X., Liu, M., Chen, G., Yuan, L., Hou, J., Zhu, S., et al. (2022). TMT-based comparative proteomic analysis of the male-sterile mutant *ms01* sheds light on sporopollenin production and pollen development in wucai (*Brassica campestris* l.). *J. Proteomics* 254, 104475. doi: 10.1016/j.jprot.2021.104475

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3 (1), 2–20. doi: 10.1093/mp/ssp106

Wang, J. W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138 (4), 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, K., Guo, Z. L., Zhou, W. T., Zhang, C., Zhang, Z. Y., Lou, Y., et al. (2018). The regulation of sporopollenin biosynthesis genes for rapid pollen wall formation. *Plant Physiol.* 178 (1), 283–294. doi: 10.1104/pp.18.00219

Wang, Y., Yang, X., Yadav, V., Mo, Y., Yang, Y., Zhang, R., et al. (2020). Analysis of differentially expressed genes and pathways associated with male sterility lines in watermelon *via* bulked segregant RNA-seq. *3 Biotech.* 10 (5), 222. doi: 10.1007/s13205-020-02208-2

Wang, Z., Li, N., Yu, Q., and Wang, H. (2021). Genome-wide characterization of salt-responsive miRNAs, circRNAs and associated ceRNA networks in tomatoes. *Int. J. Mol. Sci.* 22 (22), 12238. doi: 10.3390/ijms222212238

Wei, C., Zhang, R., Yue, Z., Yan, X., Cheng, D., Li, J., et al. (2021). The impaired biosynthetic networks in defective tapetum lead to male sterility in watermelon. *J. Proteomics*, 243, 104241. doi: 10.1016/j.jprot.2021.104241

Wei, C., Zhu, C., Yang, L., Zhao, W., Ma, R., 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. *Hortic. Res.* 6 (1), 132. doi: 10.1038/s41438-019-0213-8

Wen, M., Shen, Y., Shi, S., and Tang, T. (2012). miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. *BMC Bioinf.* 13, 140. doi: 10.1186/1471-2105-13-140

Wilson, Z. A., Morroll, S. M., Dawson, J., Swarup, R., and Tighe, P. J. (2001). The arabidopsis *MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant J.* 28 (1), 27–39. doi: 10.1046/j.1365-313x.2001.01125.x

Wu, H. J., Ma, Y. K., Chen, T., Wang, M., and Wang, X. J. (2012). PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res.* 40 (Web Server issue), W22–W28. doi: 10.1093/nar/gks554

Xing, L., Zhang, D., Li, Y., Zhao, C., Zhang, S., Shen, Y., et al. (2014). Genome-wide identification of vegetative phase transition-associated microRNAs and target predictions using degradome sequencing in *Malus hupehensis*. *BMC Genomics* 15 (1), 1125. doi: 10.1186/1471-2164-15-1125

Xu, B., Wu, R., Shi, F., Gao, C., and Wang, J. (2022). Transcriptome profiling of flower buds of male-sterile lines provides new insights into male sterility mechanism in alfalfa. *BMC Plant Biol.* 22 (1), 199. doi: 10.1186/s12870-022-03581-1

Xu, J., Ding, Z., Vizcay-Barrena, G., Shi, J., Liang, W., Yuan, Z., et al. (2014). *ABORTED MICROSPORES* acts as a master regulator of pollen wall formation in arabidopsis. *Plant Cell* 26 (4), 1544–1556. doi: 10.1105/tpc.114.122986

Xu, J., Yang, C., Yuan, Z., Zhang, D., Gondwe, M. Y., Ding, Z., et al. (2010). The *ABORTED MICROSPORES* regulatory network is required for postmeiotic male reproductive development in arabidopsis thaliana. *Plant Cell* 22 (1), 91–107. doi: 10.1105/tpc.109.071803

Yadav, V., Wang, Z., Wei, C., Amo, A., Ahmed, B., Yang, X., et al. (2020). Phenylpropanoid pathway engineering: an emerging approach towards plant defense. *Pathogens* 9 (4), 312. doi: 10.3390/pathogens9040312

Yang, Z., Yang, C., Wang, Z., Yang, Z., Chen, D., and Wu, Y. (2019). LncRNA expression profile and ceRNA analysis in tomato during flowering. *PloS One* 14 (1), e0210650. doi: 10.1371/journal.pone.0210650

Yi, L., Wang, Y., Wang, F., Song, Z., Li, J., Gong, Y., et al. (2022). Comparative transcriptome analysis reveals the molecular mechanisms underlying male sterility in autotetraploid watermelon. *J. Plant Growth Regul.* 42, 335–347. doi: 10.1007/s00344-021-10550-9

Yu, D., Li, L., Wei, H., and Yu, S. (2020). Identification and profiling of microRNAs and differentially expressed genes during anther development between a genetic male-sterile mutant and its wildtype cotton *via* high-throughput RNA sequencing. *Mol. Genet. Genomics* 295 (3), 645–660. doi: 10.1007/s00438-020-01656-y

Yuan, S., Li, Z., Yuan, N., Hu, Q., Zhou, M., Zhao, J., et al. (2020). MiR396 is involved in plant response to vernalization and flower development in *Agrostis stolonifera*. *Hortic. Res.* 7 (1), 173. doi: 10.1038/s41438-020-00394-x

Yue, Z., Ma, R., Cheng, D., Yan, X., He, Y., Wang, C., et al. (2021). Candidate gene analysis of watermelon stripe pattern locus *ClSP* ongoing recombination suppression. *Theor. Appl. Genet.* 134 (10), 3263–3277. doi: 10.1007/s00122-021-03891-2

Zhang, R., Chang, J., Li, J., Lan, G., Xuan, C., Li, H., et al. (2021). Disruption of the bHLH transcription factor *Abnormal tapetum 1* causes male sterility in watermelon. *Hortic. Res.* 8 (1), 258. doi: 10.1038/s41438-021-00695-9

Zhang, W. (2006). Regulation of arabidopsis tapetum development and function by *DYSFUNCTIONAL TAPETUM1* (*DYT1*) encoding a putative bHLH transcription factor. *Development* 133 (16), 3085–3095. doi: 10.1242/dev.02463

Zhang, Z. B., Zhu, J., Gao, J. F., Wang, C., Li, H., Li, H., et al. (2007). Transcription factor *AtMYB103* is required for anther development by regulating tapetum development, callose dissolution and exine formation in arabidopsis. *Plant J.* 52 (3), 528–538. doi: 10.1111/j.1365-313X.2007.03254.x

Zhu, J., Chen, H., Li, H., Gao, J.-F., Jiang, H., Wang, C., et al. (2008). Defective in *Tapetal development and function 1* is essential for anther development and tapetal function for microspore maturation in arabidopsis. *Plant J.* 55 (2), 266–277. doi: 10.1111/j.1365-313X.2008.03500.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 May 2023
DOI 10.3389/fpls.2023.1166008

OPEN ACCESS

EDITED BY
Jie Zhang,
Beijing Academy of Agricultural and Forestry Sciences, China

REVIEWED BY
Yong Xu,
Beijing Academy of Agriculture and Forestry Sciences, China
Hao Li,
Northwest A & F University, China

*CORRESPONDENCE
Xuanmin Dang
evergreen088@163.com
Songbi Chen
songbichen@catas.cn
Zhilong Bie
biezl@mail.hzau.edu.cn

†These authors have contributed equally to this work

RECEIVED 14 February 2023
ACCEPTED 11 April 2023
PUBLISHED 15 May 2023

CITATION
Zhan Y, Hu W, He H, Dang X, Chen S and Bie Z (2023) A major QTL identification and candidate gene analysis of watermelon fruit cracking using QTL-seq and RNA-seq. *Front. Plant Sci.* 14:1166008. doi: 10.3389/fpls.2023.1166008

A major QTL identification and candidate gene analysis of watermelon fruit cracking using QTL-seq and RNA-seq

Yuanfeng Zhan[1,2†], Wei Hu[2†], Huang He[2], Xuanmin Dang[2*], Songbi Chen[2*] and Zhilong Bie[1*]

[1]College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, China, [2]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

Fruit cracking decreases the total production and the commercial value of watermelon. The molecular mechanisms of fruit cracking are unknown. In this study, 164 recombinant inbred lines (RILs) of watermelon, derived from the crossing of the WQ1 (cracking-sensitive) and WQ2 (cracking-tolerant) lines, were sequenced using specific length amplified fragment sequencing (SLAF-seq). A high-density genetic linkage map was constructed with 3,335 markers spanning 1,322.74 cM, at an average 0.40 cM across whole-genome flanking markers. The cracking tolerance capacity (CTC), depth of fruit cracking (DFC), rind thickness (RT), and rind hardness (RH) were measured for quantitative trait locus (QTL) analysis. Of the four traits analyzed, one major QTL with high phenotypic variation (41.04%–61.37%) was detected at 76.613–76.919 cM on chromosome 2, which contained 104 annotated genes. Differential gene expression analysis with RNA sequencing (RNA-seq) data between the two parents identified 4,508 differentially expressed genes (DEGs). Comparison of the genes between the QTL region and the DEGs obtained eight coexisting genes. Quantitative real-time PCR (qRT-PCR) analysis revealed that these genes were significant differentially expressed between the two parents. These results provide new insights into the identification of QTLs or genes and marker-assisted breeding in watermelon.

KEYWORDS

fruit cracking, SLAF-seq, QTL-seq, RNA-seq, DEGs

1 Introduction

Watermelon (*Citrullus lanatus*) belongs to the cucurbit family (Cucurbitaceae) and is a popular fruit worldwide. It contains nutritional compounds such as sugar, lycopene, citrulline, arginine, and glutathione (Collins et al., 2007). In 2021, China produced 60.9 million tonnes of watermelon, accounting for about 76.3% of crops worldwide, making it

one of the top 10 watermelon-producing countries (FAO, 2021). Planting watermelon brings huge benefits to farmers in China. However, the fruit cracking of watermelon during pre- and post-harvest increases the production cost and reduces the economic value of the fruit.

Fruit cracking is a physiological disorder that occurs during fruit growth and development in many crops, such as watermelon, tomato, grape, and apple (Wang et al., 2021b). It is a complex trait associated with morphological, environmental, and genetic factors (Khadivi-Khub, 2015; Capel et al., 2017). Morphological factors, such as shape and rind thickness, affect the stability of the peel (Khadivi-Khub, 2015). Environmental factors, such mineral nutrition, endogenous hormones, water, and temperature, are associated with fruit cracking. The gibberellic acid inhibitor uniconazole P was reported to significantly reduce fruit cracking, whereas the application of excessive nitrogen fertilizer increased fruit cracking (Shimizu, 2005). Water in both the soil and the fruit may influence fruit cracking (Beyer et al., 2002; Gibert et al., 2007). Sudden moisture and temperature changes dramatically increase the fruit temperature and exacerbate fruit cracking (Simon, 2006). Regarding genetic factors, the heritability of fruit cracking in different genetic generations significantly varies (Qi et al., 2015), and different cultivars show different cracking susceptibility even under the same environmental conditions (Khadivi-Khub, 2015), indicating that genetic factors play an important role in fruit cracking.

Fruit cracking is controlled by quantitative trait loci (QTLs) (Vaidyanathan et al., 2006). A large number of QTLs related to fruit cracking have been studied in tomato, sweet cherry, and grape (Capel et al., 2017; Kunihisa et al., 2019; Crump et al., 2022; Zhang et al., 2022). However, for watermelon, only a few QTLs or genes related to fruit cracking have been mapped using bulked segregant analysis (BSA-seq) and QTL mapping (Sun et al., 2020; Yang et al., 2021; Osae et al., 2022). For example, *ClERF4* was shown to be associated with variability in fruit rind hardness (Liao et al., 2020). RNA sequencing (RNA-seq) analysis was used to screen out eight differentially expressed genes (DEGs) between cracking-resistant and cracking-susceptible parents (Jiang et al., 2019). However, the lack of experimental methods to induce cracking has made characterizing this trait a challenge (Capel et al., 2017). Thus, more QTLs related to fruit cracking must be identified.

In this study, the inbred lines WQ1 (cracking-sensitive) and WQ2 (cracking-tolerant) were crossed to generate 164 recombinant inbred lines (RILs). The RILs were then sequenced using specific length amplified fragment sequencing (SLAF-seq) and a high-density genetic map was constructed. Four fruit cracking-related traits—cracking tolerance capacity (CTC), depth of fruit cracking (DFC), rind thickness (RT), and rind hardness (RH)—were evaluated 28 days after pollination (DAP). One major QTL detected for these four traits with high phenotypic variation explained (PVE; 41.04%–61.37%) was found to be located on chromosome 2 and harbored 104 candidate genes. A comparison of the genes in the genetic region and those identified by RNA-seq revealed eight coexisting genes. Quantitative real-time PCR (qRT-PCR) analysis revealed that the eight coexisting genes were significantly differentially expressed between the two parents. Thus, these results provide new insights into mapping or cloning the QTLs or genes of fruit cracking-related traits and could be useful in marker-assisted breeding (MAS).

2 Materials and methods

2.1 Plant materials and population development

An F_8 RIL population consisting of 164 lines was generated by self-crossing the watermelon inbred lines WQ1 (female parent, cracking-sensitive) and WQ2 (male parent, cracking-tolerant). WQ1 and WQ2 were from the Tropical Crop Germplasm Research Institute, Chinese Academy of Tropical Agricultural Sciences.

2.2 Trait measurement

The F_8 lines and the two parents were planted in a greenhouse in 2021 (Danzhou, Hainan, China). All plants were grown in wide–narrow rows, with 0.45 m between plants and within a wide row of 1.0 m and a narrow row of 0.5 m. The second or third female flower per plant was artificially pollinated, and the date of pollination was recorded. Only one fruit was reserved for each plant. Field management followed normal watermelon production practices.

The RH, CTC, DFC, and RT of WQ1, WQ2, and the RILs were measured at 28 DAP. Three mature fruits per line were harvested for trait measurement. The mechanical properties of the rinds were measured using TA.XTplus Texture Analyzer (Stable Micro Systems Ltd., Godalming, UK). Three sites on the equatorial zone of each fruit were selected for RH measurement using a P/2E probe. The measurement parameters were set as described (Liao et al., 2020): the pretest speed was 1.00 m/s, the test speed was 2.00 mm/s, the posttest speed was 10 mm/s, and the distance was 15 mm. The RH value was obtained by quantifying the texture characteristic curve. CTC was measured with a knife probe (HDP/BS-B probe). Only one site on the equatorial zone of each fruit (on the reverse side of the RH measuring point) was analyzed. The measurement parameters were the same as those of the RH measurement. The breaking force when the pressure has a sudden decrease was defined as the CTC value. DFC represented the distances (in millimeters) from the contact of the knife probe to the breaking of the rind. The DFC value was calculated as the time multiplied by the test speed. RT was measured with a digital display Vernier caliper as described (Ma and Liu, 2005). For each trait, the average value of three biological replicates was calculated and taken as the phenotypic value.

To evaluate the pericarp morphology of WQ1 and WQ2 growing under normal field conditions, fruits were picked at 10 and 18 DAP. Each genotype contained three biological replicates with two fruits included. The fresh mesocarp was cut out (0.5 cm × 0.5 cm) and fixed with 5% FAA fixative (38% formaldehyde/glacial acetic acid/70% alcohol, 5:5:90, by volume), as described (Guo et al., 2020). The samples were washed twice with 50% ethanol (each time for 10 min) and dehydrated twice through an ascending ethanol

series (50%, 70%, 80%, 90%, and 100%, each time for 15 min). Subsequently, the samples were treated with 100% ethanol and xylene (3:1, 1:1, and 1:3, by volume) and 100% xylene, and with xylene and paraffin (3:1, 1:1, and 1:3, by volume) and paraffin (two times, each time at 55°C). Finally, 10-μm-thick paraffin sections were produced. A Zeiss biological microscope was used to observe and photograph the sections. Six fields of view were randomly selected for each sample. ImagePro plus 6.0 software was used to evaluate the structure of the pericarp, including the length and thickness of the epidermal cell, thickness of the exocarp, and areas of the exocarp and mesocarp cells.

2.3 DNA extraction

Young leaves from the two parents and 164 RILs were collected for genomic DNA extraction using a modified CTAB method (Saghai-Maroof et al., 1984). DNA was quantified with a NanoDrop 2000 spectrophotometer (NanoDrop, Wilmington, DE, USA) and evaluated by electrophoresis in 1.0% agarose gel. High-quality DNA samples were stored at −20°C until sequencing.

2.4 SLAF library construction and high-throughput sequencing of the RILs

The SLAF-seq strategy was used in this study (Biomarker Technologies, Beijing, China) (Sun et al., 2013). Genomic DNA from 164 RILs was digested with *Hae*III and *Hpy*166II [New England Biolabs (NEB), Ipswich, MA, USA]. Subsequently, a single nucleotide (A) was added to the 3′ end of the digested fragments using Klenow Fragment (NEB) and dATP. T4 DNA ligase was used to ligate the Duplex tag-labeled sequencing adapters to distinguish them from raw sequencing data. PCR was performed using forward (5′-AATGATACGGCGACCACCGA-3′) and reverse (5′-CAAGCAGAAGACGGCATACG-3′) primers. The PCR products were purified and pooled. The pooled samples were then separated using 2% agarose gel electrophoresis. Fragments ranging from 314 to 414 bp (with indices and adaptors) were excised and purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Paired-end sequencing (126 bp from both ends) was performed using an Illumina HiSeq 2500 System (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's instructions. To evaluate the accuracy of the SLAF libraries, the japonica rice Nipponbare (*Oryza sativa* L.) (http://rice.plantbiology.msu.edu/) was used as a control with the same process of library construction and sequencing.

2.5 SNP identification and genotyping

SLAF marker identification and genotyping were performed as described (Sun et al., 2013). Low-quality reads with quality scores <20e were filtered out. Clean reads were obtained by trimming the barcodes and terminal 5-bp positions and were then mapped onto the watermelon genome (http://cucurbitgenomics.org/ftp/genome/watermelon/97103/v2/) using BWA software (Li, 2013). Sequences mapping to the same position with >95% identity were defined as one SLAF locus (Zhang et al., 2015). The alleles of each SLAF locus were defined according to parental reads with a sequence depth of greater than fivefold; however, for each offspring, reads with a sequence depth of greater than twofold were used to define alleles. All polymorphic SLAF loci were genotyped with consistency in the parental and offspring SNP loci. Because the RIL populations were constructed using two inbred lines, segregation type aa × bb was used to genotype the SLAF markers in the RILs. The parental genotypes were aa (WQ1) and bb (WQ2), and the offspring genotypes were aa and bb (ab was removed). Polymorphic SLAF markers with the following characteristics were filtered out: 1) parental sequencing depth less than fivefold; 2) number of SNPs >3; 3) integrity filtering to screen markers covered <60% of all offspring genotypes; and 4) missing parental information on filtered sites.

2.6 Construction of the high-density genetic map

The polymorphic SLAF markers were assigned to different chromosomes by aligning with the watermelon genome. Each chromosome was considered a linkage group (LG). The modified logarithm of odds (MLOD) scores between markers were calculated to confirm the robustness of the markers for each LG. Markers with MLOD scores <3 were filtered out. The HighMap software (http://highmap.biomarker.com.cn/) was used to order the SLAF markers and correct the genotyping errors within LGs (Liu et al., 2014). Map distances were estimated using the Kosambi mapping function (Kosambi, 2016).

2.7 QTL mapping

QTLs were identified using composite interval mapping (CIM) with the R/QTL package (Broman et al., 2003). LOD values were determined based on the 1,000-permutation test. A marker with an LOD value of 3 was considered a putative QTL related to a certain trait in a genomic region.

2.8 RNA-seq analysis

For WQ1, a few fruits began to crack at 15 DAP. To identify more DEGs before and after fruit cracking, the rinds of WQ1 and WQ2 fruits that were artificially pollinated at 10 and 18 DAP (before and after fruit cracking, respectively) were collected for RNA extraction. Total RNA was extracted using the TransZol Up Plus RNA Kit, qualified by electrophoresis with the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), and purified using the RNAClean XP Kit and RNase-Free DNase Set. RNA-seq libraries were constructed after ribosomal RNA (rRNA) removal, fragmentation, first- and second-strand complementary DNA (cDNA) synthesis, end repair, A tailing to

the 3′ end, ligation of adapters, and enrichment by PCR amplification. The concentration and the size of the constructed library were detected using Qubit 2.0 Fluorometer and Agilent 4200, respectively. Paired-end (150 bp) sequencing was performed with the Illumina HiSeq X Ten System following the manufacturer's protocol.

Adaptors and low-quality bases of the raw reads were removed with the FASTX-Toolkit. Short reads of <25 bp were discarded. Clean reads were mapped to the watermelon_97103_v2 genome by HISAT2 (Kim et al., 2015), allowing four mismatches, and unique matches were used to calculate the gene read number and fragments per kilobase of transcript per million fragments mapped (FPKM) (Trapnell et al., 2010).

2.9 Differentially expressed genes

Two groups (WQ1 *vs*. WQ2 at 10 DAP and WQ1 *vs*. WQ2 at 18 DAP) were used to detect the DEGs. DEGs were screened out using the DESeq2 package in R (Love et al., 2014). A fold change ≥2 and a *q*-value ≤0.05 were set as the cutoff values.

2.10 Expression analysis of the candidate genes

Total RNA was extracted from the rind of WQ1 and WQ2 at 10 DAP using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Thereafter, cDNA was synthesized from 3 μg gDNA Eraser-treated (Takara, San Jose, CA, USA) RNA. cDNA corresponding to 30 ng of the total RNA was used as the template for each SYBR Green PCR reaction, which utilized the ABI ViiA 7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). qRT-PCR was performed with three biological and three technical replicates for each candidate gene. *Cla97C05G094190* (*GAPDH*) was used as the internal control (Itoh et al., 2023). The primers for qRT-PCR were designed using Primer 5 (Lalitha, 2000) and are listed in Supplementary Table S1.

3 Results

3.1 Phenotypic analysis of the parents and RILs

WQ1 is a fruit cracking-sensitive inbred line that has a rate of 77.8% fruit cracking in the field (Figures 1A, C), while WQ2 is an inbred line that is fruit cracking-tolerant (Figures 1B, D). The RH and CTC values of WQ2 at 10 and 18 DAP were significantly higher than those of WQ1 (Figures 1I, J). The CTC values of WQ2 were 1.85 and 2.85 times higher than those of WQ1 at 10 and 18 DAP, respectively (Figures 1I, J). The RH values of both WQ1 and WQ2 increased after artificial pollination. The CTC value of WQ1 decreased at 18 DAP, whereas that of WQ2 increased.

The length of the epidermal cells in WQ1 was significantly higher than that in WQ2 at each stage (at 10 and 18 DAP), but the thickness of the epidermal cells was only significantly different between WQ1 and WQ2 at 18 DAP. The epidermal cells of WQ2 were almost square, and their size showed little change during fruit development. The length-to-thickness ratio varied from 1.10 to 1.21. In contrast, the epidermal cells of WQ1 were rectangular, and the cells were elongated with fruit development. The length-to-thickness ratio changed from 1.57 to 2.11 (Figures 1E–H and Supplementary Table S2).

The exocarp of the fruit in WQ2 had 9–12 layers of cells, and its thickness changed from 119.01 to 137.69 μm at a different stage. The exocarp cells were small and closely arranged and short oval or short polygonal in shape. In contrast, the exocarp of the fruit in WQ1 was thinner and consisted of four to six layers of cells, with its thickness changing from 76.29 to 78.17 μm at a different stage. The exocarp cell area of WQ1 cells was 1.75–3.18 times larger than that of WQ2 cells (Figures 1E–H and Supplementary Table S2).

The mesocarp cell area of WQ1 was significantly larger than that of WQ2. The mesocarp cells of WQ2 and WQ1 gradually became larger with the development of the fruit. However, the mesocarp cells of WQ2 changed little during fruit development, with the cell area increasing by only 133.79 μm^2 from 10 to 18 DAP. However, the mesocarp cells of WQ1 changed greatly during fruit development. The area was 2,238.93 μm^2 at 10 DAP, which increased to 3,925 μm^2 at 18 DAP (Figures 1E–H and Supplementary Table S2).

For QTL mapping, the CTC, DFC, RT, and RH of the RILs and WQ1 and WQ2 were measured at 28 DAP. Significant differences were observed in the fruit cracking-related traits between the two parents and the RILs (Table 1). The average CTC, DFC, RH, and RT values of the RILs were 14.23 kg (1.58–37.20 kg), 8.48 mm (1.51–18.17 mm), 55.95 kg (29.40–78.45 kg), and 6.37 mm (2.09–13.74 mm), respectively. All four traits exhibited super-parent segregation and normal distribution in the RILs (Figure 2). The correlation between the four traits showed significant differences (Supplementary Table S3), indicating that these traits were closely related to fruit cracking.

3.2 Construction of the genetic map

After the clean reads were mapped to the watermelon reference genome, 197,639 and 198,988 SNP markers were obtained from WQ1 and WQ2, respectively. A total of 494,293 SNP markers were developed between the parents and the RILs (Supplementary Table S4). Finally, 175,062 SNP markers were obtained by filtering out those markers that had no polymorphism between the parents or the SNPs that were missing in the offspring or distributing partial separation (Supplementary Table S5). The filtered SNPs were classified into seven types (i.e., aa × bb, ab × cc, cc × ab, ef × eg, hk × hk, lm × ll, and nn × np) (Supplementary Figure S1). Because the population used in this study was the RIL population constructed by two diploid inbred lines, the aa × bb type (132,967 SNP markers) was reserved for genetic analysis. After fine filtering, 4,857 SNP markers were finally obtained for the construction of the genetic map.

The MLOD values between the two markers were calculated for the 4,857 SNP markers. The highest MLOD values between the

10.3389/fpls.2023.1166008

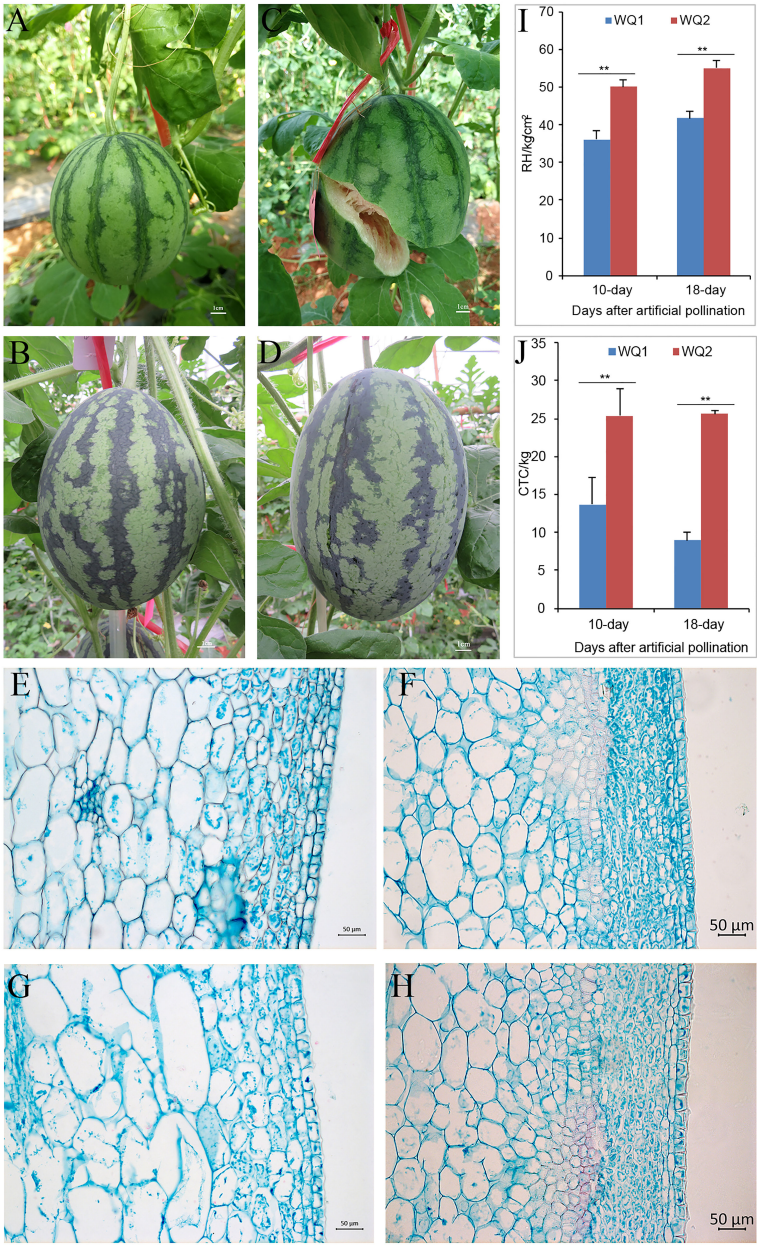


FIGURE 1
Phenotypes of parents. **(A, B)** WQ1 and WQ2 at 10 days after pollination (DAP). **(C, D)** WQ1 and WQ2 at 18 DAP. **(E, F)** Pericarp morphology of WQ1 and WQ2 at 10 DAP. **(G, H)** Pericarp morphology of WQ1 and WQ2 at 18 DAP. **(I, J)** Rind hardness (RH) and cracking tolerance capacity (CTC) of fruits at 10 and 18 DAP. $^{**}p < 0.05$.

TABLE 1 Fruit-cracking traits related to the parents and RILs.

Trait	Parents		RILs		VC (%)	Skewness	Kurtosis
	WQ1	WQ2	Average	Range			
CTC (kg)	4.52	24.64**	14.23	1.58–37.20	52.07	0.579	−0.421
DFC (mm)	3.11	14.81**	8.48	1.51–18.17	41.58	0.399	−0.050
RH (kg/cm^2)	45.92	60.58**	55.95	29.40–78.45	18.65	−0.107	−0.549
RT (mm)	3.18	8.62**	6.37	2.09–13.74	39.08	0.284	−0.665

Asterisks mark significant differences according to Student's t-test.
CTC, cracking tolerance capacity; DFC, depth of fruit cracking; RH, rind hardness; RIL, recombinant inbred line; RT, rind thickness; VC, variable coefficient.
$^{**}p < 0.01$.

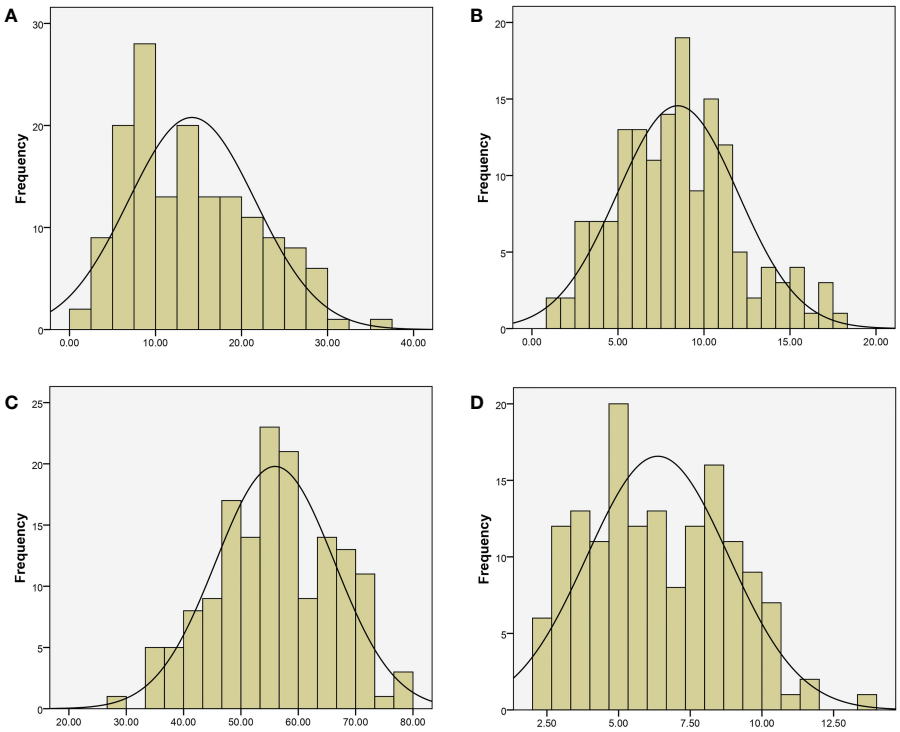


FIGURE 2
Phenotypic distribution of recombinant inbred lines (RILs). **(A)** Cracking tolerance capacity (CTC, in kilograms). **(B)** Depth of fruit cracking (DFC, in millimeters). **(C)** Rind hardness (RH in kilograms per square centimeter); **(D)** rind thickness (RT in millimeters).

markers were classified into the same LG. A total of 3,335 SNP markers were obtained, accounting for 68.66% of the total markers. A high-density genetic map including 11 chromosomes was constructed using HighMap software (Figure 3). The number of SNP markers in each chromosome ranged from 113 to 538. A genetic map with 1,322.74 cM length was developed, ranging from 63.48 to 144.02 cM among all the chromosomes. The average genetic distance between the SNP markers across the chromosomes was 0.4 cM. The average genetic distances on chromosomes 2 (0.73 cM) and 5 (0.26 cM) were the largest and smallest across all the chromosomes, respectively. The ratio of gaps ≤5 cM in each chromosome ranged from 96.43% to 99.26% (Supplementary Table S6). Generally, the higher the ratio of gaps <5 cM to the total number of gaps, the more uniform the map. The maximum distance on different chromosomes ranged from 6.24 to 18.58, of which the maximum distance between chromosome 5 was the smallest and that of chromosome 3 was the largest (Supplementary Table S7).

3.3 Detection of QTLs associated with fruit cracking

In total, two, one, one, and two QTLs were detected for CTC, DFC, RT, and RH, respectively, when the LOD threshold was set to 3 (Table 2). Of all the detected QTLs, one locus with high PVE (41.04%–61.37%) was detected for each of the four traits, which was located at 76.613–76.919 cM on chromosome 2 (Figure 4). The physical location at this interval was 31.804–32.805 Mb. A total of 104 genes were predicted in this interval. All the additive effects of

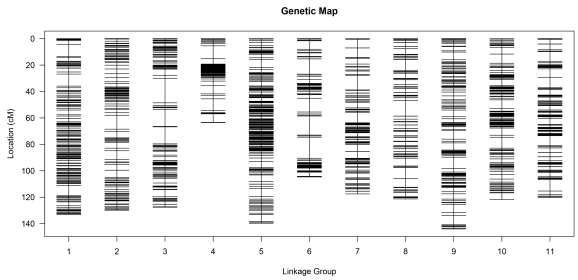


FIGURE 3
High-density genetic map.

10.3389/fpls.2023.1166008

this QTL were negative, indicating that the WQ2 allele of this locus is the desirable one for fruit cracking resistance. Another minor QTL, which just exceeded the LOD thresholds for CTC and RT, was detected on chromosome 7 at the same position, with PVE of 8.66% and 11.2% (Table 2).

3.4 RNA-seq analysis

For each repeat, >57,616,037 clean reads were obtained, with matching rates ranging from 80.17% to 86.04%. The proportion of reads matching multiple positions was between 1.05% and 1.53% (<10%), indicating that the sequencing results of all samples were of high quality for subsequent analysis (Supplementary Table S7).

In total, 2,551 DEGs were detected between WQ1 and WQ2 at 10 DAP, with 1,342 upregulated and 1,209 downregulated DEGs (Supplementary Figure S2A and Supplementary Data Sheet 1). In addition, 3,642 DEGs were detected between WQ1 and WQ2 at 18 DAP, with 1,963 DEGs that were upregulated and 1,679 that were downregulated (Supplementary Figure 1B and Supplementary Data Sheet 2). In the DEGs detected in the two groups, 1,685 genes were co-detected. A total of 866 and 1,957 unique DEGs were detected at 10 and 18 DAP, respectively (Supplementary Figure S3).

3.5 Gene ontology analysis of DEGs

Gene Ontology (GO) analysis was conducted for the functional classification of the DEGs. In total, 2,290 (89.7%) DEGs were assigned to three GO classes—biological process, cellular component, and molecular function—with 2,289 DEGs classified into 49 functional groups in the WQ1 *vs*. WQ2 at 10 DAP group (Figure 5A). In biological process, the DEGs were mainly classified into cellular process (n = 1,175), material metabolism (n = 1,100), and single-organism process (n = 1,064). In cellular component, the DEGs were mainly classified into cells (n = 1,898), cell parts (n = 1,898), and organelles (n = 1,373). In molecular function, the DEGs were mainly classified into protein binding (n = 1,047) and enzyme catalysis activity (n = 899).

For the WQ1 *vs*. WQ2 at 18 DAP group, 3,230 (88.7%) DEGs were assigned to three GO classes, with 2,290 DEGs classified into 51 functional groups (Figure 5B). In biological process, the DEGs were mainly classified into cellular process (n = 1,655), material metabolism (n = 1,588), and single-organism process (n = 1,471). In cellular component, the DEGs were mainly classified into cells (n = 2,764), cell parts (n = 2,762), and organelles (n = 1,964). In molecular function, the DEGs were mainly classified into protein binding (n = 1,480) and enzyme catalysis activity (n = 1,355).

3.6 KEGG analysis of DEGs

The Kyoto Encyclopedia of Genes and Genomes (KEGG) categories involved cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. A total of 309 DEGs were assigned to 92 KEGG pathways in the WQ1 *vs*. WQ2 at 10 DAP group (Supplementary Data Sheet 3). The significantly enriched pathways included phenylpropanoid biosynthesis; flavonoid biosynthesis; secondary metabolite biosynthesis; phenylalanine metabolism; and biosynthesis of stilbene, heptane, and gingerol (Figure 5C).

In total, 450 DEGs were assigned to 107 KEGG pathways in the WQ1 *vs*. WQ2 at 18 DAP group (Supplementary Data Sheet 4). The significantly enriched pathways included pentose and gluconate interconversion; phenylalanine biosynthesis; cutin, flavin, and wax biosynthesis; and linoleic acid metabolism (Figure 5D).

3.7 Selection of candidate genes

The comparison of the DEGs detected through RNA-seq and the genes on the major QTL interval obtained eight coexisting genes, of which seven were upregulated and one was downregulated (Table 3). The eight coexisting genes included four transcription factors (*Cla97C02G043800*, *Cla97C02G044050*, *Cla97C02G044440*, and *Cla97C02G044520*) and four other annotated genes (*Cla97C02G043690*, *Cla97C02G043750*, *Cla97C02G043850*, and *Cla97C02G044100*) (Figure 6A and Table 3). The coexisting genes were verified by qRT-PCR. Among the seven upregulated genes, the expression of *Cla97C02G044520* and *Cla97C02G043690* between WQ1 and WQ2 showed the largest and the smallest difference, respectively. For the only downregulated gene, *Cla97C02G044050*, the expression of WQ1 was 20 times of WQ2 (Figure 6B). The expression of all detected genes reached significant levels.

TABLE 2 QTLs detected of the CTC, DFC, RT, and RH.

QTL	Max LOD score	Group	Genetic position (cM)	Additive	PVE (%)
qCTC-1	31.7	2	76.613–76.919	−5.765	60.83
qCTC-2	3.1	7	80.127	2.174	8.66
qDFC	17.8	2	76.613–76.919	−2.253	41.04
qRH	18.1	2	76.613–76.919	−6.677	41.6
qRT-1	32.1	2	76.613–76.919	−1.937	61.37
qRT-2	4	7	77.907–80.127	0.829	11.2

CTC, cracking tolerance capacity; DFC, depth of fruit cracking; LOD, limit of detection; PVE, phenotypic variation explained; QTL, quantitative trait locus; RH, rind hardness; RIL, recombinant inbred line; RT, rind thickness.

4 Discussion

Flesh fruits, such as apple, sweet cherry, grape, and tomato, can crack or split during growth and development, causing severe economic loss (Khadivi-Khub, 2015). QTLs related to fruit cracking have been mapped for some fruits with different populations (Capel et al., 2017; Kunihisa et al., 2019; Crump et al., 2022; Zhang et al., 2022). However, only a few studies have focused on the fruit cracking of watermelon (Jiang et al., 2019; Liao et al., 2020; Sun et al., 2020).

Accurate measurement of phenotypes is important for QTL mapping. Only a few QTLs have been identified by counting the number of cracking fruits, calculating the cracked fruit rate, or evaluating the melon cracking capacity with high levels of irrigation (Huang et al., 1999; Qi et al., 2015; Capel et al., 2017). However, no QTL with a large effect was identified at the same time, which is unfavorable for gene identification or applications. RH has been reported as a reliable indicator of cracking resistance capacity (Liao et al., 2020). The CTC, DFC, RT, and RH values showed high correlations with each other in the correlation analysis (Supplementary Table S3). One major locus was mapped to the same region by quantitative trait locus sequencing (QTL-seq), indicating that, in addition to RH, the phenotype of CTC, DFC, and RT can be used for the QTL mapping of fruit cracking.

Compared with traditional molecular markers—such as random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), cleaved amplified polymorphic sequence (CAPS), or simple sequence repeat (SSR) markers—a lot of markers can be detected through high-throughput sequencing. In this study, a genetic map of 1,322.74 cM length was constructed, which had a higher density than the map developed using molecular markers (Hashizume et al., 2003; Cheng et al., 2016; Yang et al., 2021; Osae et al., 2022). However, the maximum gap on chromosomes 3 and 6 was relatively large (Supplementary Table S6), indicating that the sequencing fold may have been increased. High-density genetic maps in watermelons have been rarely reported.

QTL-seq and RNA-seq have been used for QTL identification in many plants (Park et al., 2019; Wen et al., 2019; Lei et al., 2020; Xue et al., 2022). However, for some agronomic traits, it is almost impossible to identify the QTLs using RIL or F_2 populations due to fewer recombination sites and complex genetic backgrounds. Although a minor QTL was identified on chromosome 7, it had little contribution to the fruit cracking capacity between the two parents (Table 2). Therefore, the QTL located on chromosome 2 can be considered a quality trait locus, which can be fine-mapped using $F_{2:3}$ populations or introgression lines.

During fruit development, the pericarp structures of WQ2 and WQ1 were significantly different (Figures 1E–H). The RH and CTC values of WQ2 were consistently significantly higher than those of WQ1. The thickness of the pericarp and exocarp of WQ2 was significantly higher than that of WQ1, and the lengths and areas of the epicarp and mesocarp cells of WQ1 were significantly greater than those of WQ2 (Supplementary Table S2). In later fruit development, the length of the epidermal cells, the area of the epicarp cells, and the area of the mesocarp cells of WQ1 were 2.91,

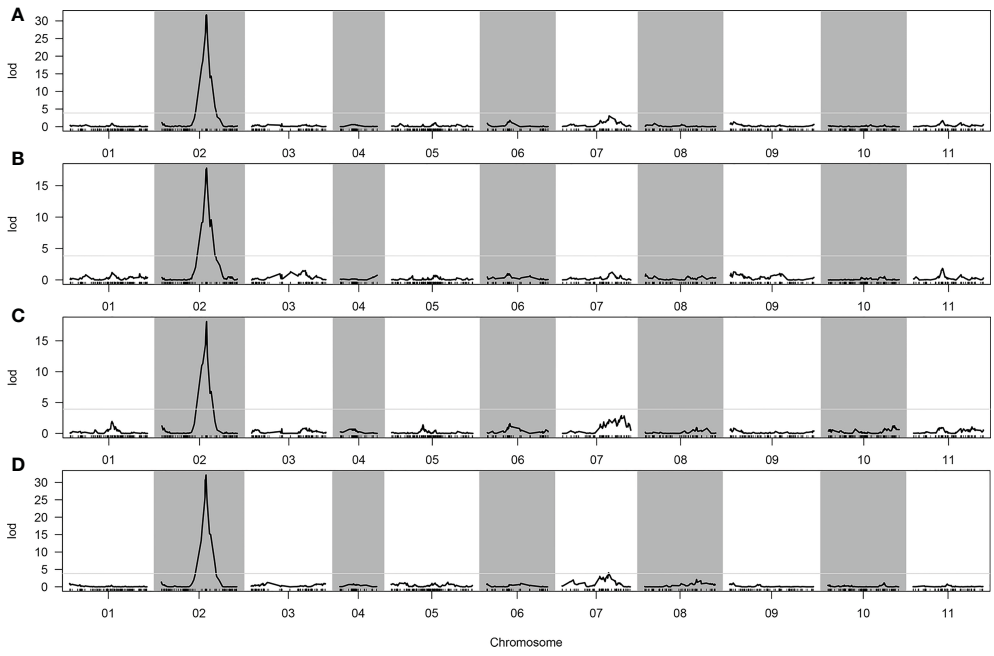


FIGURE 4
Quantitative trait loci (QTLs) detected in the whole genome. **(A)** Cracking tolerance capacity (CTC). **(B)** Depth of fruit cracking (DFC). **(C)** Rind hardness (RH). **(D)** Rind thickness (RT).

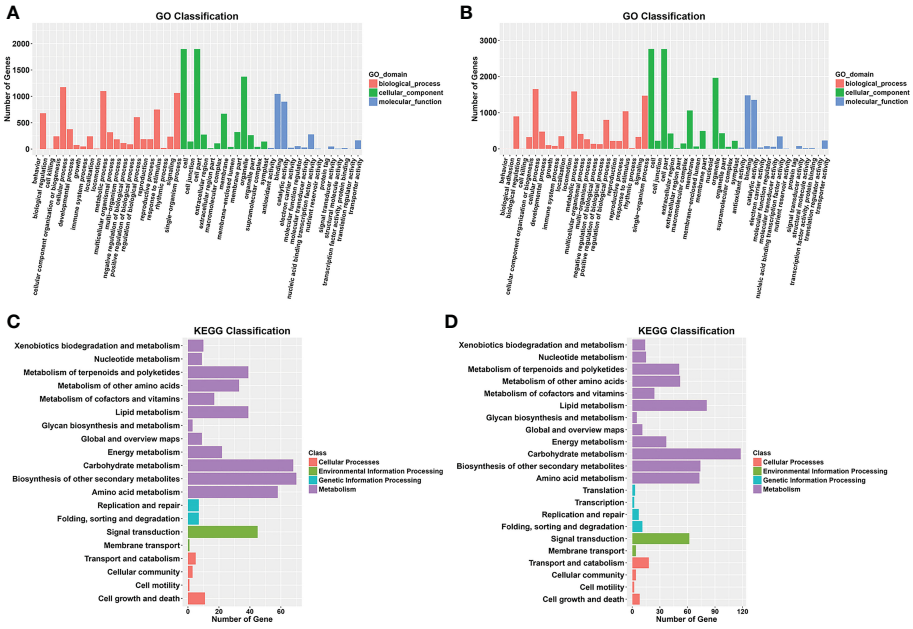


FIGURE 5
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of the differentially expressed genes (DEGs) from the two comparison groups. **(A, B)** GO enrichment analysis between WQ1 and WQ2 at 10 and 18 days after pollination (DAP). **(C, D)** KEGG enrichment analysis between WQ1 and WQ2 10 at and 18 DAP.

3.28, and 5.69 times larger than those of WQ2, respectively. The RH and CTC values were significantly positively correlated with pericarp and exocarp thickness, but significantly negatively correlated with epidermal cell length and the exocarp and mesocarp cell areas. Therefore, RH and CTC can be used to measure the cracking resistance of watermelon fruit, which is more appropriate at the later stage of fruit development.

In total, 866 and 1,957 unique DEGs were obtained at 10 and 18 DAP, respectively (Figure 6A), indicating that, with fruit development, the gene expression was changed. The changed DEGs may be responsible for fruit cracking. DEGs changing during fruit development may play a key role in the difference between watermelon rind and fruit cracking. GO and KEGG enrichment analyses of the DEGs before and after cracking (at 10 and 18 DAP) revealed that lignin catabolism, glucuronoxylan synthesis and metabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis, secondary metabolite biosynthesis, plant-type secondary cell wall biogenesis, and regulation of jasmonic acid-mediated signaling pathways were significantly enriched. The change of DEGs may play a vital role in the formation of the rind tissue structure of watermelon during fruit development. Of the eight coexisting candidate genes both in the QTL interval and DEGs, four were transcription factors. Transcription factors are involved in the regulation of floral and fruit development and gametophyte cell division (Lai et al., 2020). Many transcription factors have been detected in fruit cracking through RNA-seq (Jiang et al., 2019; Wang et al., 2021a). This may provide new insights into the selection of candidate genes.

TABLE 3 Coexisting genes detected by comparing the QTL-seq and RNA-seq.

Gene ID	Description	P/S difference
Cla97C02G043690	Reticulon-like protein	UP
Cla97C02G043750	Seed biotin-containing protein SBP65	UP
Cla97C02G043800	Transcription factor TCP4-like	UP
Cla97C02G043850	Nucleobase-ascorbate transporter 1	UP
Cla97C02G044050	Heat stress transcription factor A-4c-like	DOWN
Cla97C02G044100	Serine/threonine protein phosphatase 7 long form	UP
Cla97C02G044440	Zinc finger protein CONSTANS	UP
Cla97C02G044520	MYB transcription factor 58.1	UP

UP, upregulated; DOWN, downregulated.

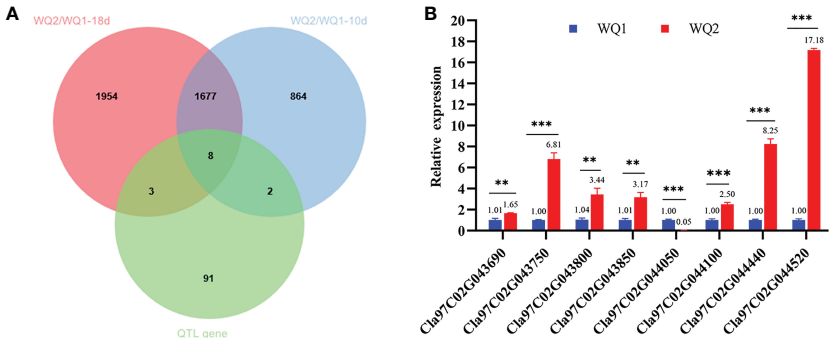


FIGURE 6
Coexisting genes in quantitative trait locus sequencing (QTL-seq) and RNA-seq and the candidate genes identified by quantitative real-time PCR (qRT-PCR). **(A)** Venn diagram of the differentially expressed genes at 10 and 18 days after pollination (DAP) and the genes within the QTL intervals. **(B)** Relative expression of the candidate genes between the two parents. *M*, maternal (WQ1); *P*, paternal (WQ2). $^{**}p < 0.01$, $^{**}p < 0.001$.

In future studies, our group aims to construct $F_{2:3}$ populations or introgression lines for the fine-mapping of major QTLs. In addition, genetic transformation of the candidate genes will be performed. The results of this study may provide a useful locus for MAS and gene function studies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, SRR23013875–SRR23014040.

Author contributions

YZ, SC, and ZB conceived the research. YZ, WH, HH, and XD performed the experiments and analyzed the data. WH and YZ drafted the manuscript. XD, SC, and ZB revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Hainan Provincial Natural Science Foundation of China (no.321RC647), Hainan Province Science and Technology Special Fund (no. ZDYF2021XDNY290), the Social Public-Interest Scientific Institution Reform Fund (no. PZS2022019), and Hainan Province Science and Technology Special Fund (no. ZDYF2022XDNY332).

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.2023.1166008/full#supplementary-material

References

Beyer, M., Peschel, S., Knoche, M., and Knoürgen, M. (2002). Studies on water transport through the sweet cherry fruit surface: IV. regions of preferential uptake. *HortScience* 37 (4), 637–641. doi: 10.21273/HORTSCI.37.4.637

Broman, K. W., Wu, H., Sen, Ś, and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7), 889–890. doi: 10.1093/bioinformatics/btg112

Capel, C., Yuste-Lisbona, F. J., López-Casado, G., Angosto, T., Cuartero, J., Lozano, R., et al. (2017). Multi-environment QTL mapping reveals genetic architecture of fruit cracking in a tomato RIL solanum lycopersicum× s. pimpinellifolium population. *Theor. Appl. Genet.* 130, 213–222. doi: 10.1007/s00122-016-2809-9

Cheng, Y., Luan, F., Wang, X., Gao, P., Zhu, Z., Liu, S., et al. (2016). Construction of a genetic linkage map of watermelon (Citrullus lanatus) using CAPS and SSR markers and QTL analysis for fruit quality traits. *Sci. Hortic.* 202, 25–31. doi: 10.1016/j.scienta.2016.01.004

Collins, J. K., Wu, G., Perkins-Veazie, P., Spears, K., Claypool, P. L., Baker, R. A., et al. (2007). Watermelon consumption increases plasma arginine concentrations in adults. *Nutrition* 23 (3), 261–266. doi: 10.1016/j.nut.2007.01.005

Crump, W. W., Peace, C., Zhang, Z., and McCord, P. (2022). Detection of breeding-relevant fruit cracking and fruit firmness quantitative trait loci in sweet cherry *via* pedigree-based and genome-wide association approaches. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.823250/full

FAO (2021). *Top 10 country production of watermelon*. Available at: https://www.fao.org/faostat/en/#rankings/countries_by_commodity.

Gibert, C., Chadœuf, J., Vercambre, G., Génard, M., and Lescourret, F. (2007). Cuticular cracking on nectarine fruit surface: spatial distribution and development in relation to irrigation and thinning. *J. Am. Soc. Hortic. Sci.* 132 (5), 583–591. doi: 10.21273/JASHS.132.5.583

Guo, Y., Zhang, T., Zhong, J., Ba, T., Xu, T., Zhang, Q., et al. (2020). Identification of the volatile compounds and observation of the glandular trichomes in opisthopappus taihangensis and four species of chrysanthemum. *Plants* 9 (7), 855. doi: 10.3390/plants9070855

Hashizume, T., Shimamoto, I., and Hirai, M. (2003). Construction of a linkage map and QTL analysis of horticultural traits for watermelon [Citrullus lanatus (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. *Theor. Appl. Genet.* 106, 779–785. doi: 10.1007/s00122-002-1030-1

Huang, X., Wang, H., Gao, F., and Huang, H. (1999). A comparative study of the pericarp of litchi cultivars susceptible and resistant to fruit cracking. *J. Hortic. Sci. Biotechnol.* 74 (3), 351–354. doi: 10.1080/14620316.1999.11511120

Itoh, T., Muramatsu, M., Miyazono, D., Koketsu, M., Fujita, S., and Hashizume, T. (2023). Phenolic glycosides citrulluside h and citrulluside T isolated from young watermelon (Citrullus lanatus) fruit have beneficial effects against cutibacterium acnes-induced skin inflammation. *Nat. Prod. Commun.* 18 (1), 1934578X221143202. doi: 10.1177/1934578X221143202

Jiang, H., Tian, H., Yan, C., Jia, L., Wang, Y., Wang, M., et al. (2019). RNA-Seq analysis of watermelon (Citrullus lanatus) to identify genes involved in fruit cracking. *Sci. Hortic.* 248, 248–255. doi: 10.1016/j.scienta.2019.01.005

Khadivi-Khub, A. (2015). Physiological and genetic factors influencing fruit cracking. *Acta Physiol. Plant* 37 (1), 1718. doi: 10.1007/s11738-014-1718-2

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

Kosambi, D. D. (2016). "The estimation of map distances from recombination values," in *DD kosambi* (New Delhi, India: Springer), 125–130. doi: 10.1007/978-81-322-3676-4_16

Kunihisa, M., Takita, Y., Yamaguchi, N., Okada, H., Sato, M., Komori, S., et al. (2019). The use of a fertile doubled haploid apple line for QTL analysis of fruit traits. *Breed. Sci.* 69 (3), 410–419. doi: 10.1270/jsbbs.18197

Lai, X., Chahtane, H., Martin-Arevalillo, R., Zubieta, C., and Parcy, F. (2020). Contrasted evolutionary trajectories of plant transcription factors. *Curr. Opin. Plant Biol.* 54, 101–107. doi: 10.1016/j.pbi.2020.03.002

Lalitha, S. (2000). Primer premier 5. *Biotech. Software. Internet Report.: Comput. Software. J. Sci.* 1 (6), 270–272. doi: 10.1089/152791600459894

Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., Wang, J., 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, 1–14. doi: 10.1186/s12284-020-00416-1

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv. preprint. arXiv:1303.3997*. doi: 10.48550/arXiv.1303.3997

Liao, N., Hu, Z., Li, Y., Hao, J., Chen, S., Xue, Q., et al. (2020). Ethylene-responsive factor 4 is associated with the desirable rind hardness trait conferring cracking resistance in fresh fruits of watermelon. *Plant Biotechnol. J.* 18 (4), 1066–1077. doi: 10.1111/pbi.13276

Liu, D., Ma, C., Hong, W., Huang, L., Liu, M., Liu, H., et al. (2014). Construction and analysis of high-density linkage map using high-throughput sequencing data. *PloS One* 9 (6), e98855. doi: 10.1371/journal.pone.0098855

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 (12), 1–21. doi: 10.1186/s13059-014-0550-8

Ma, S., and Liu, J. (2005). *Watermelon germplasm resources description specification and data standard* (Beijing: China Agricultural Press).

Osae, B. A., Amanullah, S., Liu, H., Liu, S., Saroj, A., Zhang, C., et al. (2022). CAPS marker-base genetic linkage mapping and QTL analysis for watermelon ovary, fruit and seed-related traits. *Euphytica* 218 (4), 39. doi: 10.1007/s10681-022-02990-5

Park, M., Lee, J.-H., Han, K., Jang, S., Han, J., Lim, J.-H., et al. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of capsicum chinense revealed using QTL-seq and RNA-seq. *Theor. Appl. Genet.* 132, 515–529. doi: 10.1007/s00122-018-3238-8

Qi, Z., Li, J., Raza, M. A., Zou, X., Cao, L., Rao, L., et al. (2015). Inheritance of fruit cracking resistance of melon (Cucumis melo l.) fitting e-0 genetic model using major gene plus polygene inheritance analysis. *Sci. Hortic.* 189, 168–174. doi: 10.1016/j.scienta.2015.04.004

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 81 (24), 8014–8018. doi: 10.1073/pnas.81.24.8014

Shimizu, T. (2005). Effects of GA inhibitor on the fruit cracking of melon (Cucumis melo l.). *Hortic. Res.* 4, 89–93. doi: 10.2503/hrj.4.89

Simon, G. (2006). Review on rain induced fruit cracking of sweet cherries (Prunus avium l.), its causes and the possibilities of prevention. *Int. J. Hortic. Sci.* 12 (3), 27–35. doi: 10.31421/IJHS/12/3/654

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PloS One* 8 (3), e58700. doi: 10.1371/journal.pone.0058700

Sun, L., Zhang, Y., Cui, H., Zhang, L., Sha, T., Wang, C., et al. (2020). Linkage mapping and comparative transcriptome analysis of firmness in watermelon (Citrullus lanatus). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00831

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28 (5), 511–515. doi: 10.1038/nbt.1621

Vaidyanathan, S., Harrigan, G. G., and Goodacre, R. (2006). *Metabolome analyses:: strategies for systems biology* (New York, United States: Springer Science & Business Media).

Wang, Y., Guo, L., Zhao, X., Zhao, Y., Hao, Z., Luo, H., et al. (2021b). Advances in mechanisms and omics pertaining to fruit cracking in horticultural plants. *Agronomy* 11 (6), 1045. doi: 10.3390/agronomy11061045

Wang, J., Wu, X., Tang, Y., Li, J. G., and Zhao, M. (2021a). RNA-Seq provides new insights into the molecular events involved in "Ball-skin versus bladder effect" on fruit cracking in litchi. *Int. J. Mol. Sci.* 22 (1), 454. doi: 10.3390/ijms22010454

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., et al. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC Plant Biol.* 19, 1–17. doi: 10.1186/s12870-019-2008-3

Xue, Y., Gao, H., Liu, X., Tang, X., Cao, D., Luan, X., et al. (2022). QTL mapping of palmitic acid content using specific-locus amplified fragment sequencing (SLAF-seq) genotyping in soybeans (Glycine max l.). *Int. J. Mol. Sci.* 23 (19), 11273. doi: 10.3390/ijms231911273

Yang, T., Amanullah, S., Pan, J., Chen, G., Liu, S., Ma, S., et al. (2021). Identification of putative genetic regions for watermelon rind hardness and related traits by BSA-seq and QTL mapping. *Euphytica* 217, 1–18. doi: 10.1007/s10681-020-02758-9

Zhang, C., Cui, L., Liu, C., Fan, X., and Fang, J. (2022). Mining candidate genes of grape berry cracking based on high density genetic map. *Hortic. Plant J.* doi: 10.1016/j.hpj.2022.10.004

Zhang, J., Zhang, Q., Cheng, T., Yang, W., Pan, H., Zhong, J., et al. (2015). High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (Prunus mume sieb. et zucc). *DNA Res.* 22 (3), 183–191. doi: 10.1093/dnares/dsv003

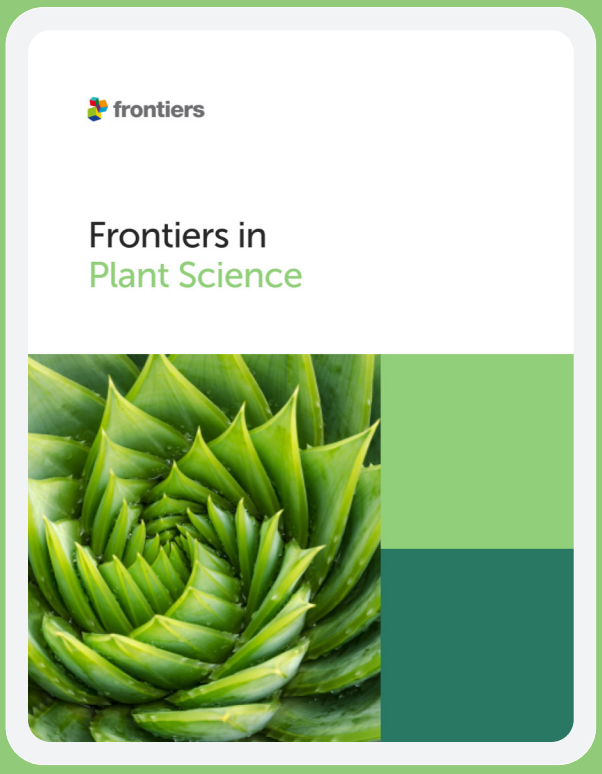
Frontiers in
Plant Science

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