TRANSLATIONAL RESEARCH FOR CUCURBIT MOLECULAR BREEDING: TRAITS, MARKERS, AND GENES

EDITED BY: Yiqun Weng, Amnon Levi, Jordi Garcia-Mas and Feishi Luan

frontiers Research Topics

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: researchtopics@frontiersin.org

TRANSLATIONAL RESEARCH FOR CUCURBIT MOLECULAR BREEDING: TRAITS, MARKERS, AND GENES

Topic Editors:
Yiqun Weng, University of Wisconsin-Madison, United States
Amnon Levi, United States Department of Agriculture, United States
Jordi Garcia-Mas, Institute of Agrifood Research and Technology (IRTA), Spain
Feishi Luan, Northeast Agricultural University, China

Table of Contents

frontiers
in Plant Science

Editorial: Translational Research for Cucurbit Molecular Breeding: Traits, Markers, and Genes

Yiqun Weng[1]*, Jordi Garcia-Mas[2,3], Amnon Levi[4] and Feishi Luan[5]

[1] United States Department of Agriculture, Agriculture Research Service, Vegetable Crops Research Unit, Horticulture Department, University of Wisconsin, Madison, WI, United States, [2] Center for Research in Agricultural Genomics (CRAG) Consejo Superior de Investigaciones Científicas - Institut de Recerca i Tecnologia Agroalimentàries - Universitat Autònoma de Barcelona - Universitat de Barcelona, Barcelona, Spain, [3] Genomics and Biotecnology Program, Institut de Recerca i Tecnologia Agroalimentáries (IRTA), Barcelona, Spain, [4] US Vegetable Laboratory, United States Department of Agriculture, Agriculture Research Service, Charleston, SC, United States, [5] Key Laboratory of Biology and Genetic Improvement of Horticulture Crops (Northeast Region), Ministry of Agriculture and Rural Affairs, Northeast Agricultural University, Harbin, China

Editorial on the Research Topic

Translational Research for Cucurbit Molecular Breeding: Traits, Markers, and Genes

INTRODUCTION

Cucurbits (family Cucurbitaceae) are economically important vegetable crops. Major cucurbits growing globally include cucumber, melon, watermelon, and squash/pumpkin. Other cucurbits like bitter melon, bottle gourd, winter melon, and luffa are popular in many Asian and African countries. The last decade has witnessed a rapid development of genetic and genomics resources including draft genome assemblies, and high-density genetic maps in a dozen cucurbit crops, making it possible to accelerate translational research for cucurbit breeding. This Research Topic is a collection of 21 Original Research articles or Reviews highlighting the achievements and future directions in cucurbit translational research. These articles cover a variety of topics ranging from improvement of the cucurbit genome assemblies to identification and molecular mapping of horticulturally important genes or QTL for horticultural traits, and the use of such knowledge in marker-assisted selection for cucurbit improvement. Major findings from these investigations are summarized below.

IMPROVEMENT OF MELON DRAFT GENOME ASSEMBLY

Draft genomes of nearly a dozen cucurbit crops are now available (https://www.cucurbitgenomics.org/) and are constantly being revised using new technologies and experimental data. Castanera et al. reported a further improvement of the melon reference genome using PacBio single-molecule real time (SMRT) sequencing, which produced melon genome version v4.0 (DHL92). The melon reference genome v3.5.1 (Garcia-Mas et al., 2012) was obtained using 454 technology and the genome assembly was further improved using optical mapping to produce v3.6.1 (Ruggieri et al., 2018). However, v3.6.1 still contained 19% of gaps and more than 40 Mb unassigned sequences, probably missing complex repeat regions. DHL92 melon assembly v4.0 had an increase of the melon genome pseudomolecule size by 40 Mb with 90% of the v3.5.1 gaps being filled, and the completeness was improved mainly in non-genic regions. Specifically, 40% more full-length LTR retrotransposons were identified in the new assembly, mainly located in centromeric and pericentromeric regions, and a burst of these repetitive elements was found to occur less than two

million years ago. Some of these elements are polymorphic among melon varieties and sit in the upstream regions of genes; however, their potential regulatory roles are unknown.

QTL MAPPING OF DISEASE RESISTANCES IN CUCURBITS

Cucurbits are susceptible to a number of common pathogens. Identification of disease resistances and associated molecular markers is often a priority in most cucurbit breeding programs. Three articles reported molecular mapping of host resistances against viral pathogens. Pascual et al. describe a general strategy for virus resistance in melon, conferred by the *vacuolar protein sorting 41* (*CmVPS41*) gene, which is responsible for the restriction of the virus to the bundle sheath cells and preventing phloem entry. VPS41 is a protein involved in intracellular trafficking of proteins and vesicles from late Golgi to the vacuole. A previous study (Giner et al., 2017) found that the *cmv1* locus in melon accession PI 161375 for resistance against the Cucumber Mosaic Virus (CMV) was due to a mutation in *CmVPS41*. Pascual et al. sequenced the *CmVPS41* locus in 52 melon accessions and identified 16 new haplotypes, 12 protein variants, and nine new resistant accessions. Only two non-synonymous mutations, L348R carried by PI 161375 and G85E, resulted in CMV resistance. In both cases, the virus was able to replicate and move from cell to cell but was unable to reach the phloem. The authors suggest that resistance to viral entrance to the phloem seems to be a general strategy in melon. This opens up new possibilities for using different breeding approaches based on genes, which control different steps of the viral cycle to promote resistances to different viruses.

The cucumber vein yellowing virus (CVYV) causes severe yield losses in cucurbit crops across Mediterranean countries. Pujol et al. identified *CsCvy-1*, a monogenic locus with incomplete dominance for CVYV resistance, in a long Dutch cucumber, which can explain more than 80% of the variance. A BSA (bulked segregant analysis)-Seq and an additional linkage analysis placed *CsCvy-1* in a 625 kb region of cucumber Chromosome 5 (Chr5). There are 24 annotated genes in this region including two *RNA-dependent RNA polymerase* (*RDR1a* and *RDR1b*) genes as potential candidates for *CsCvy-1*. RDRs are critical players in RNA silencing pathways involved in the process of amplification of double-stranded RNAs that activate gene silencing after nuclease processing.

Sáez et al. conducted a QTL analysis for resistance to the Tomato leaf curl New Delhi virus (ToLCNDV) in two pumpkin (*Cucurbita moschata*) accessions (PI 419083 and PI 604506). Allelism tests indicated that both lines carried the same monogenic, recessive locus. Consistent with this, QTL mapping identified a major-effect QTL for ToLCNDV resistance on Chr8, which was also confirmed in the interspecific *C. moschata* × *C. pepo* segregating populations, but the resistance level seemed to be affected by the *C. pepo* genetic background. A comparative analysis indicated that this major-effect QTL was located in a syntenic region on melon Chr11 which was previously shown to harbor a locus for ToLCNDV resistance.

Two other articles reported microRNA (miRNA)-regulated resistance to target leaf spot (TLS, causal agent *Corynespora cassiicola*) and QTL mapping of Fusarium wilt (FW, caused by *Fusarium oxysporum* f. sp. *cucumerinum*) resistances in cucumber. Wang X. et al. studied the association of four miRNAs, miR164d, miR396b, Novel-miR1, and Novel-miR7, with TLS resistance. The four miRNAs inhibit their target genes *NAC transcription factor*, *APE* (for anthranilate phosphoribosyltransferase), *4CL* (coding for 4-coumarate: CoA ligase), and *PAL* (coding for phenylalanine ammonia-lyase), respectively. Transient expression of the four miRNAs and the corresponding four target genes in cucumber cotyledons, resulted in reduced and enhanced TLS resistance, respectively. Overexpression of *4CL, PAL*, and *novel-miRNA1/7* down-regulated lignin synthesis suggesting that the two miRNAs are associated with TLS resistance through regulating flavonoid biosynthesis. In another article, Dong J. et al. conducted QTL mapping of FW resistance in the cucumber line "Rijiecheng." A major-effect QTL, *fw2.1* was detected in a 1.91-Mb-long region of cucumber Chr2. Further linkage analysis narrowed the QTL into a 600 kb region containing ∼80 annotated genes. RNA-Seq revealed five genes in this region which were up-regulated in the FW resistant parental line. This QTL was adjacent to two FW resistance QTL previously detected in this region.

QTL MAPPING FOR ABIOTIC STRESS TOLERANCES IN CUCUMBER AND MELON

Two articles in this Research Topic report on QTL mapping of abiotic stress tolerances. Dong S. et al. conducted QTL mapping for low temperature tolerance (LTT) in cucumber seedlings. Phenotyping was performed in replicated trials using an $F_{2:3}$ population and a low-temperature injury index (LTII) that was of quantitative inheritance. Three QTL, *qLTT5.1*, *qLTT6.1*, and *qLTT6.2* were detected, with the major-effect QTL *qLTT6.2* accounting for ∼25% of the observed phenotypic variance. *In silico* BSA and qPCR (quantitative real-time PCR) in a re-sequenced cucumber core collection identified a 42-kb region of the *qLTT6.2* locus with *Csa6G445210* (coding for an auxin response factor), and *Csa6G445230* (coding for an ethylene-responsive transmembrane protein) being the most possible candidates for this QTL.

In another study in melon, Branham et al. report QTL mapping for tolerance to sulfur phytotoxicity and the development of markers useful for melon breeding programs. Sulfur is widely used in cucurbit cultivation to control powdery mildew, although it is known that several accessions can develop a phytotoxicity reaction leading to plant death. The use of the MR-1 ×AY RIL (recombinant inbred line) population that segregates for sulfur tolerance allowed the identification of one major (*qSulf-1* on Chr1) and two minor (*qSulf-8* on Chr8, and *qSulf-12* on Chr12) QTL. The authors also provided a set of KASP (Competitive allele specific PCR$^©$) markers tightly linked to *qSulf-1*, which can immediately be incorporated into melon breeding programs.

MOLECULAR MAPPING OF FRUIT YIELD AND QUALITY-RELATED TRAITS

For all cucurbits, fruit is the main target for breeding improvement. Important attributes affecting cucurbit fruit yield, and quality may include fruit size/shape, fruit external (rind/skin color, stripe pattern, fruit spines, warts, and texture), and internal (flavor/taste, nutritional content, flesh color) qualities. In this Research Topic, seven articles described studies on fruit quality-related traits in cucumber, watermelon and bottle gourd. Lycopene content and flesh color are important fruit nutritional and appearance attributes for watermelon. Wang C. et al. conducted linkage analysis in a red × pale yellow flesh F_2 population and found that the red flesh color was controlled by a recessive gene, which was located in a 392 kb region on Chr4. Further linkage and association analyses narrowed the candidate gene down to a ~41 kb region. Additional evidence supported the *lycopene β-cyclase* (*LCYB, Cla005011*) gene as the candidate that regulates flesh color changes at the protein level. Cleaved amplified polymorphic sequence (CAPS) markers were developed to differentiate red vs. yellow flesh colors in watermelon genotypes that could be used in marker-assisted breeding.

Bitter fruit is undesirable for consumption in bottle gourd (*Lagenaria siceraria*). Wu et al. phenotyped an $F_{2:3}$ population with a trained panel of sensory analysis and found two complementary genes underlying fruit bitterness in bottle gourd, which was consistent with the QTL mapping result in this population. The two QTL, *QBt.1* and *QBt.2*, were located in a region of 1.6-Mbp and 1.9-Mbp on Chr6, and Chr7, respectively, which were validated in an advanced bulked segregant analysis (A-BSA). Sequence-based comparative analysis suggested that the causal genes underlying *QBt.1* and *QBt.2* may not be direct orthologs of the reported bitterness genes in cucumber, melon and watermelon.

Among the four articles on cucurbit external fruit quality attributes, Liu et al. investigated the inheritance of rind (skin) colors in watermelon. BSA-Seq and fine mapping with a yellow × green F_2 population mapped the *yellow rind* (*Clyr*) locus in a 91.42 kb region on Chr4. The authors also conducted comparative transcriptomic analysis in the two parental lines at different fruit development stages and identified genes and pathways related to rind pigment metabolism. Rett-Cadman et al. studied developmental changes and natural variation of cucumber fruit surface properties. They first conducted a microscopic investigation of the fruit surface in two cucumber lines (CL9930 and Gy14) at different development stages and found significant differences in cuticle thickness, depth of wax intercalation between epidermal cells, epidermal cell size and shape, and number and size of lipid droplets in the two lines. QTL mapping in the Gy14 × CL9930 RIL population revealed strong QTL for epidermal cell height, cuticle thickness, intercalation depth, and diameter of lipid droplets co-localized on cucumber Chr1. Fine mapping combined with gene expression profiling suggested a small number of candidate genes underlying these traits. Tissue specificity, developmental analysis of expression,

allelic diversity, and gene function analyses identified the transcription factor *CsSHINE1/CsWIN1* (*CsV3_1g030200*) as a source of natural variation for cucumber fruit epidermal traits. Pericarp wax on cucumber fruit skin may affect consumer preference and marketability. Grafting of cucumber onto pumpkin rootstock (*Cucurbita moschata*) often produces glossy fruits. Zhang J. et al. conducted transcriptome profiling, genome-wide DNA methylation sequencing, and wax metabolite analysis in the pericarps of self-rooted vs. grafted cucumbers, and found that the AP2/ERF-type transcription factor gene *CsWIN1* (*Csa3G878210*) was methylated and up-regulated in grafted compared to self-rooted scions. The increased expression of *CsWIN1* was positively correlated with expression of several key wax biosynthesis genes and the accumulation of wax esters in grafted cucumber pericarp. The glossier appearance of grafted pericarp seemed to be the result of higher wax ester content and higher integration of small trichomes in the pericarp. It is interesting to see that the two different members of the AP2/ERF-type transcription factor gene family control different aspects of fruit epidermal features.

The fruit spine is also an important external fruit quality trait. The MYB transcription factor gene *CsTRY* (*Csa5G139610*) is a negative regulator for fruit spine initiation. Zhang L. et al. found that in plants overexpressing *CsTRY*, the expression of key genes in the flavonoid synthesis pathway was repressed, suggesting that CsTRY not only regulates the development of fruit spines, but also functions in the synthesis of flavonoids, acting as the repressor of anthocyanin synthesis.

Cucumber is consumed when still immature. Over ripe fruit affects its market value. Cheng et al. isolated and characterized a MADS-box gene, *Cucumis sativus SHATTERPROOF* (*CsSHP*) in cucumber that may participate in fruit maturation through the ABA pathway and floral organ specification *via* interaction with CsSEPs (SEPALLATA).

QTL FOR SEED-RELATED TRAITS

In many countries, seed are also the end products for watermelon and pumpkin/squash (*Cucurbita* spp.) production. There is considerable variation in seed color, size, and shape in these crops. Four articles reported on a genetic analysis and development of molecular markers for seed color and size in watermelon and pumpkin, and a comparative analysis QTL for seed size variation among cucurbits. Paudel et al. examined the previously established four-gene model (*R*, *T1*, *W*, and *D*) explaining seed coat color variation in watermelon. QTL-seq and genotyping-by-sequencing (GBS) analyses mapped *R*, *T1*, *W*, and *D* on Chr3, 5, 6, and 8, respectively. They also developed KASP assays and SNP markers useful for incorporating each of these gene loci in watermelon breeding programs. In another study, Li B. et al. conducted a genetic mapping analysis and identified a *polyphenol oxidase* (*PPO*) gene (*Cla019481*) as the candidate for the *ClCS1* locus controlling the black seed color in watermelon. The black color was proposed to be due to a frameshift mutation

in this *PPO* gene that is involved in melanin oxidation during melanin biosynthesis.

In pumpkin (*C. maxima*), based on a high-density SNP-based genetic map, Wang Y. et al. identified 10 QTL for seed width (SW), seed length (SL), and hundred-seed weight (HSW). Based on gene annotation and non-synonymous SNPs in the major SL and SW-associated regions, two genes coding for a VQ motif and an E3 ubiquitin-protein ligase were proposed to be the potential candidates influencing SL, while an F-box and leucine-rich repeat (LRR) domain-containing protein is the potential regulator for SW in *C. maxima*.

Guo et al. conducted a literature review on seed size (SS) QTL identified in watermelon, pumpkin/squash, cucumber, and melon, and inferred consensus SS QTL based on their physical positions in respective draft genomes. Among them, four from watermelon (*ClSS2.2, ClSS6.1, ClSS6.2,* and *ClSS8.2*), two from cucumber (*CsSS4.1* and *CsSS5.1*), and one from melon (*CmSS11.1*) were a major-effect, stable QTL for seed size and weight, which were located in syntenic regions across different genomes, suggesting possible structural, and functional conservation of some important genes for seed size control in cucurbit crops.

PLANT ARCHITECTURE IN WATERMELON AND SEX EXPRESSION IN CUCURBITS

Plant architecture is important for crop improvement. There is great interest by breeders and growers in the development of dwarf watermelon varieties useful for high plant density cultivation and high yield production. Zhu et al. conducted BSA-Seq of F_2 plants derived from a cross between a normal-height and dwarf watermelon inbred lines. They identified a gene (*Cla010337*) coding for an ATP-binding cassette transporter (ABC transporter) protein with two SNPs associated with the dwarfness. A derived CAPS (dCAPS) marker that co-segregates with the dwarf trait was validated using the F_2 population and a germplasm collection of 165 watermelon accessions. This dCAPS marker can be used in a marker-assisted selection (MAS) for the development of dwarf watermelon cultivars.

In cucurbits, sex determination is critically associated with fruit earliness, yield, and quality. Li D et al. reviewed this important topic on how sex determination genes and their interactions determine sex in cucurbit crops. Six sex morphs could be found based on the distribution of female, male, or bisexual flowers on a plant. So far, five orthologous genes involved in sex determination have been cloned, and their various combinations and expression patterns can explain all the identified sex types in cucurbits. The phytohormone ethylene plays a critical role in sex expression in cucurbits. Two ethylene signaling components, CsERF110 and CsERF31, have recently been identified, which may help in exploring the ethylene signaling-mediated interactions among sex-related genes. The authors proposed a nomenclature to name genes and sex morphs across different cucurbit crops, which is at present rather confusing.

CONCLUSIONS

This Research Topic provides readers with updates of translational research in cucurbit crops, many updates of which are not only of importance for cucurbit breeding, but also elucidate basic concepts of cucurbit development and response to biotic and abiotic stressors. The editors thank all contributors to this Research Topic.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, B., González, V. M., et al. (2012). The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109

Giner, A., Pascual, L., Bourgeois, M., Gyetvai, G., Rios, P., Picó, B., et al. (2017). A mutation in the melon vacuolar protein sorting 41 prevents systemic infection of *Cucumber mosaic virus*. *Sci. Rep.* 7:10471. doi: 10.1038/s41598-017-10783-3

Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., et al. (2018). An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. *Sci. Rep.* 8:8088. doi: 10.1038/s41598-018-26416-2

Functional Identification of *Corynespora cassiicola*-Responsive miRNAs and Their Targets in Cucumber

Xiangyu Wang[1,2], Guangchao Yu[1,2], Junyue Zhao[1], Na Cui[1], Yang Yu[1] and Haiyan Fan[1,3]*

[1] College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, China, [2] College of Horticulture, Shenyang Agricultural University, Shenyang, China, [3] Key Laboratory of Protected Horticulture of Ministry of Education, Shenyang Agricultural University, Shenyang, China

Target leaf spot (TLS), which is caused by *Corynespora cassiicola* (*C. cassiicola*), is one of the most important diseases in cucumber (*Cucumis sativus* L.). Our previous research identified several *C. cassiicola*-responsive miRNAs in cucumber by high-throughput sequencing, including two known miRNAs and two novel miRNAs. The target genes of these miRNAs were related to secondary metabolism. In this study, we verified the interaction between these miRNAs and target genes by histochemical staining and fluorescence quantitative assays of GUS. We transiently expressed the candidate miRNAs and target genes in cucumber cotyledons to investigate the resistance to *C. cassiicola*. Transient expression of miR164d, miR396b, Novel-miR1, and Novel-miR7 in cucumber resulted in decreased resistance to *C. cassiicola*, while transient expression of *NAC* (inhibited by miR164d), *APE* (inhibited by miR396b), *4CL* (inhibited by Novel-miR1), and *PAL* (inhibited by Novel-miR7) led to enhanced resistance to *C. cassiicola*. In addition, overexpression of *4CL* and *PAL* downregulated lignin synthesis, and overexpression of Novel-miR1 and Novel-miR7 also downregulated lignin synthesis, indicating that the regulation of *4CL* and *PAL* by Novel-miR1 and Novel-miR7 could affect lignin content. The tobacco rattle virus (TRV) induced short tandem target mimic (STTM)-miRNA silencing vector was successfully constructed, and target miRNAs were successfully silenced. The identification of disease resistance and lignin content showed that silencing candidate miRNAs could improve cucumber resistance to *C. cassiicola*.

Keywords: cucumber, *Corynespora cassiicola*, microRNA, transient transformation, short tandem target mimic, lignin

INTRODUCTION

Cucumber (*Cucumis sativus* L.) target leaf spot (TLS) is caused by *Corynespora cassiicola* (*C. cassiicola*), which is an obligate oomycete pathogen (Teramoto et al., 2011; Li et al., 2012). TLS affects a wide range of cucumbers worldwide, including those in China (Li et al., 2012; Yang et al., 2012), the United States (Ishii et al., 2007), Japan (Miyamoto et al., 2010), and South Korea (Kwon et al., 2003). TLS is a foliar disease that can occur throughout the growing period of cucumber, but it is more serious in the middle and late stages of growth than in other stages (Wen et al., 2015).

There are abundant variations in the size and structure of the *C. cassiicola* conidia, as well as the color of the colonies. These variations are reflected not only in different colonies of the same strain but also among strains of different geographical origins and different hosts (Nghia et al., 2008; Qi et al., 2011). Therefore, it is important to study the molecular mechanisms of cucumber resistance to *C. cassiicola* and find new resistance gene resources.

Plant microRNAs (miRNAs) are a class of small non-coding single stranded RNAs encoded by endogenous genes that are mainly involved in gene expression and regulation at the post-transcriptional level (Waterhouse and Hellens, 2015). In plants, miRNAs can function by interacting with target genes, mainly through degradation and inhibition of target mRNAs (Slack et al., 2000; Khraiwesh et al., 2012). Plants are often subjected to various environmental stresses during their growth. These stresses can result in accumulation of various substances and induce plant-related gene expression and metabolic pathways to strengthen plant resistance. Recently, many genes encoding stress-related proteins have been discovered, and miRNAs have been shown to have important regulatory roles in the expression of these genes (Sanan-Mishra et al., 2009; Baxter et al., 2014; Liu et al., 2015; Zhang and Wang, 2015). Many plant miRNAs can be induced after infection by pathogens, and these miRNAs can participate in plant disease resistance by interacting with their targets (Thiebaut et al., 2015). miR393 was the first miRNA reported to participate in the interaction between plants and pathogens (Navarro et al., 2006). In *Arabidopsis thaliana*, miR824, miR843, miR852, miR166, miR156, and miR159 could respond to *Pseudomonas syringae* (Zhang et al., 2011). Overexpression of miR482b in tomato reduced resistance to *Phytophthora infestans* by inhibiting expression of *NBS-LRR*, while silencing miR482b increased resistance (Jiang et al., 2018). In cucumber, several studies have identified many miRNAs that are related to growth and stress responses (Martínez et al., 2011; Mao et al., 2012; Li et al., 2014; Jin and Wu, 2015; Burkhardt and Day, 2016), but there are currently no reports of miRNAs related to resistance to *C. cassiicola*.

After infection by pathogens, plants will produce secondary metabolites to resist pathogen invasion (Pateraki and Kanellis, 2010; Pusztahelyi et al., 2015). In plants, there are two main types of secondary metabolites that affect disease resistance. One is the inherent substances of plants, such as lignin, callose, and keratin. These substances can reinforce cell walls and prevent pathogens from damaging plant tissues (Zhao and Dixon, 2011). The other group includes alkaloids, terpenes, and phenols that are induced by pathogens. These substances have a direct bactericidal effect (Agrawal and Weber, 2015). Lignin is an important component of the cell wall, with a complex structure and induced properties. Increasing the lignin content can enhance the ability of plant cells to resist penetration and dissolution and inhibit the spread of pathogenic bacteria in plants (Zhao and Dixon, 2011; Li et al., 2015). *Phenylalanine ammonia-lyase* (*PAL*) and *4-coumarate: CoA ligase* (*4CL*) are the two key genes involved in phenylpropanoid synthesis. They can promote the synthesis of lignin to enhance plant disease resistance (Kim and Hwang, 2014; Li et al., 2015).

In our previous study, high-throughput sequencing was performed to investigate the differentially expressed miRNAs in cucumber inoculated with *C. cassiicola*, including two known miRNAs (miR164d and miR396b) and two novel miRNAs (Novel-miR1 and Novel-miR7) (Wang et al., 2018). Based on the analyses of target genes function, we believe that these miRNAs play important roles in the interaction between cucumber and *C. cassiicola*. For further elucidation of cucumber miRNAs, it is necessary to effectively and accurately demonstrate the interaction of miRNAs and target genes. In this study, candidate miRNAs and target genes were transiently expressed in tobacco, and the interaction between miRNAs and target genes was determined by analysis of GUS histochemical staining and fluorescence quantification. Meanwhile, the candidate miRNAs and target genes were transiently expressed in the cucumber cotyledons to investigate the resistance of the transgenic cucumber to *C. cassiicola*. We analyzed the mechanism of cucumber resistance to *C. cassiicola* based on the regulation of miRNAs, which provided a theoretical reference for further stable genetic transformation and breeding research of plants.

Due to the short size of miRNAs and the redundancy of family functions, traditional gene silencing methods are not suitable for miRNA research. Target mimics (TMs) can block the inhibition of target genes by miRNAs, thereby inhibiting the regulatory function of miRNAs (Franco-Zorrilla et al., 2007; Meng et al., 2012). Tang et al. (2012) discovered a new miRNA silencing regulatory mechanism, short tandem target mimic (STTM), which has high silencing efficiency and can be widely used in the functional studies of miRNAs. The STTM is composed of two TMs and a 48 nt linkage sequence. There are three nucleotides forming a bulge between the 10th and 11th nucleotides of the TM on both sides of the STTM, and thus, the binding sites can capture miRNAs without being cleaved by them. Compared to TM, STTM has a better inhibitory effect on miRNAs (Yan et al., 2012). Virus-induced gene silencing (VIGS) can induce plant endogenous gene silencing and alter the plant phenotype to explore gene function (Sha et al., 2014). Tobacco rattle virus (TRV) is an RNA virus that can infect a variety of plants. TRV-based vectors have mild infection symptoms and high gene silencing efficiency (MacFarlane et al., 1999). TRV-based vectors have been widely used in VIGS to inhibit gene expression in a variety of plants (Bachan and Dinesh-Kumar, 2012; Huang et al., 2012; Zhou et al., 2012). However, virus-based miRNA silencing (VBMS) has not been reported in cucumber. In this study, a TRV-based VBMS system that can effectively inhibit the activity of endogenous miRNAs in cucumber for a certain period of time was developed and may provide a strategy for further analysis of the resistance mechanism of cucumber to *C. cassiicola*.

MATERIALS AND METHODS

Plant Materials

Nicotiana benthamiana (*N. benthamiana*) were planted in a greenhouse at 25°C under 16:8 light/dark cycles for 20 days for agroinfiltration. The cucumber variety used in the experiments

was Xintaimici, which was planted in a greenhouse at 28°C under 16:8 light/dark cycles for 10 days for agroinfiltration.

DNA and RNA Extraction and cDNA Synthesis

The DNA of the cucumber leaves was extracted using a Plant Genomic DNA Kit (Tiangen, Beijing, China). Total RNA was extracted by using a RNAprep Pure Plant Kit (Tiangen, Beijing, China) and synthesized into cDNA using a QuantScript RT Kit (Tiangen, Beijing, China). First-strand cDNA synthesis corresponding to miRNAs was performed using a miRcute miRNA First-Strand cDNA Synthesis Kit (Tiangen, Beijing, China).

Construction of Transient Expression Vectors of Candidate miRNAs and Their Targets

Based on the precursor sequences of miR164d, miR396b, Novel-miR1, and Novel-miR7 (**Supplementary Table S1**), primers were designed according to the In-Fusion principles. Similarly, based on the *NAC*, *APE*, *PAL*, and *4CL* open reading frame sequences (**Supplementary Table S1**), primers were also designed by the In-Fusion principles. The primers were synthesized by GENEWIZ (Suzhou, China) and are listed in **Supplementary Table S2**. The target genes were annotated via BLAST against the cucumber genome database[1]. The mRNA raw data were deposited in the NCBI Sequence Read Archive (SRA) with the accession number SRP117262; the small RNA raw data were deposited in the NCBI Sequence Read Archive (SRA) with the accession number SRP117230.

pRI-101 AN (TaKaRa, Dalian, China) is a binary vector for plant transformation that can efficiently express exogenous genes. The plant expression vectors pRI-101 AN and pRI-101 AN-GUS were constructed to verify the interaction between miRNAs and targets. For cloning cDNA into the vector, an InFusion HD cloning kit (Clontech, CA, United States) was used. All constructed vectors were confirmed by sequencing before transformation into *Agrobacterium tumefaciens* (*A. tumefaciens*) strain EHA105.

Co-transformation of Candidate miRNAs and Targets in *N. benthamiana*

To verify the interaction between candidate miRNAs and target genes, we transiently expressed constructed vectors in *N. benthamiana* leaves by *A. tumefaciens* transformation. The agrobacteria carrying constructs were diluted to OD_{600} = 0.5 with suspension buffer (10 mM MES, 10 mM $MgCl_2$ and 200 μM acetosyringone). pRI-101 AN-miRNA and pRI-101 AN-GUS-target genes were mixed in equal volumes to test the cleavage function of tae-miR408. Normal tobacco (Control) and tobacco injected with pRI-101 AN-GUS were used as controls. After 48 h of infiltration, GUS staining and fluorescence quantitative analysis were performed as described by Bradford (1976) and

[1]https://phytozome.jgi.doe.gov/pz/portal.html

Jefferson et al. (1987). All experiments were performed with three biological repeats.

Transient Expression in Cucumber Cotyledons and Disease Resistance Assays

Transient expression in cucumber cotyledons was performed as described by Shang et al. (2014). The agrobacteria carrying constructs were diluted to OD_{600} = 0.4 with suspension buffer (10 mM MES, 10 mM $MgCl_2$ and 200 μM acetosyringone) for cotyledon infiltration. We overexpressed miRNA and target genes to verify disease resistance. Leaves were collected 48 h after agroinfiltration to analyze the expression of miRNAs and target genes by qRT-PCR.

For disease resistance assays, inoculation of *C. cassiicola* was performed by pipetting multiple 10 μL droplet spore suspensions (2×10^5 sporangia/mL) onto cucumber cotyledons. The inoculated samples were kept at 100% relative humidity for 5 days, and disease was assessed by measuring lesion size and quantifying fungal biomass by qRT-PCR quantification of *C. cassiicola Actin*. The primers used were CoActin-F and CoActin-R (**Supplementary Table S2**). All experiments were performed with three biological repeats. Normal cucumber (Control) and cucumber injected with pRI-101 AN were used as controls. All experiments were performed with three biological repeats.

Construction of the TRV-Induced Silencing Vector

The sequences of STTM-miR164d, STTM-miR396b, STTM-Novel-miR1, and STTM-Novel-miR7 used in this study were synthesized by GENEWIZ (Suzhou, China). The STTM primers were designed according to the In-Fusion principles and are listed in **Supplementary Table S2**. The pTRV vector can be efficiently expressed in plants. For cloning STTM into the vector, an InFusion HD cloning kit (Clontech, CA, United States) was used. All constructed vectors were confirmed by sequencing before transformation into *A. tumefaciens* strain EHA105.

Virus-Based MicroRNA Silencing (VBMS)

Cucumber cotyledons grown for approximately 10 days were used for VBMS, and the suspension method of *Agrobacterium* was the same as described previously. The mixed suspension containing pTRV1 and pTRV2-STTM was injected into the cucumber cotyledons by a sterile syringe. Cotyledons at 7 days post-injection were used for *C. cassiicola* inoculation and disease resistance assays.

Quantitative Real-Time RT-PCR Assay

qRT-PCR analyses of genes were conducted using a SuperReal PreMix Plus Kit (SYBR Green) (Tiangen, Beijing, China), and the cucumber *Actin* gene was used as the internal control. qRT-PCR analyses of miRNAs were performed using the miRcute miRNA qRT-PCR Detection Kit (SYBR Green) (Tiangen, Beijing, China), and the U6 snRNA was used as a reference to normalize the data. All qRT-PCR assays were performed on a LightCycler 480 system

(Roche, CA, United States), and the primers used are listed in **Supplementary Table S2**. The relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), and the standard deviation was calculated with three biological repeats.

Determination of Lignin in Cucumber Cotyledons

The determination of lignin content was performed as described by Morrison (1972) with slight modifications. The lignin content is defined as the absorbance at 280 nm per gram of fresh weight. The 1 g cucumber cotyledon sample was homogenized with 95% ethanol and centrifuged at 5000 rpm for 5 min, and the sediments were washed three times with 95% ethanol. After the samples were washed twice with a mixture of ethanol and n-hexane (1:2, v/v), sediments were collected by centrifugation. The dried samples were dissolved in 2 mL of bromoacetyl and glacial acetic acid (1:3, v/v) solution. After a 30 min water bath at 70°C for 30 min, 0.9 mL 2 mM NaOH was added to stop the reaction. Then, 2 mL glacial acetic acid and 0.1 mL 7.5 M hydroxylamine hydrochloride were added to the sample, and the sample was diluted to 5 mL with glacial acetic acid. After centrifugation at 5000 rpm for 5 min, the supernatants were collected to determine the absorbance at 280 nm.

Statistical Analysis

All data are the mean (±SD) of three biological replicates. Statistical analysis was carried out by one-way analysis of variance (ANOVA) using the IBM SPSS Statistics 22 software, and the significant differences were determined by Duncan's multiple range test ($P < 0.05$) and indicated in alphabetical notation.

RESULTS

Validation of the Interaction Between Candidate miRNAs and Target Genes

Our previous research found that secondary metabolism plays an important role in cucumber in response to *C. cassiicola* infection (Wang et al., 2018). Based on previous study, two known miRNAs and two novel miRNAs with their targets were selected, as they may be involved in enhanced *C. cassiicola* resistance in cucumber (**Table 1**). In plants, miRNAs regulate their target gene mainly by recognizing a specific site on the target gene mRNA and binding to it to form a silencing complex, thereby inhibiting translation of the target gene mRNA. Candidate miRNAs and target gene binding sites are shown in **Figure 1**.

TABLE 1 | Candidate miRNAs and their targets.

MicroRNA	Targets ID	Targets description
miR164d	Cucsa.040380	NAC domain containing protein (NAC)
miR396b	Cucsa.098530	Anthranilate phosphoribosyltransferase (APE)
Novel-miR1	Cucsa.261210	4-coumarate: CoA ligase (4CL)
Novel-miR7	Cucsa.124480	Phenylalanine ammonia-lyase (PAL)

To verify the interaction, we co-transformed candidate miRNAs and target genes into tobacco leaves. Using the vector pRI-101 AN-GUS containing the *GUS* gene, we tested the inhibitory effect of miRNAs on target genes with GUS histochemical staining and fluorescence quantification. Normal tobacco (Control) and tobacco inoculated with pRI-101 AN-GUS were used as controls. The phenotypes of GUS histochemical staining are shown in **Figure 2A**. The GUS phenotypes were not observed in tobacco leaves inoculated with the recombination vector pRI-101 AN-miRNA (miR164d, miR396b, Novel-miR1, and Novel-miR7) and normal tobacco (Control). The GUS phenotypes were observed in tobacco leaves inoculated with pRI-101 AN-GUS, while leaves inoculated with pRI-101 AN-GUS-target gene (*NAC*, *APE*, *4CL*, and *PAL*), in which the target gene was fused upstream of the *GUS* gene, showed similar phenotypes. However, GUS phenotypes were markedly reduced in leaves co-transformed with pRI-101 AN-miRNA (miR164d, miR396b, Novel-miR1, and Novel-miR7) and pRI-101 AN-GUS-target gene (*NAC*, *APE*, *4CL*, and *PAL*).

The GUS protein activity in leaves inoculated with different recombinant vectors was measured by fluorescence quantitative assays (**Figure 2B**). GUS fluorescence was not detected in normal tobacco and leaves inoculated with pRI-101 AN-miRNA (miR164d, miR396b, Novel-miR1, and Novel-miR7). There were increased GUS fluorescence values in leaves inoculated with pRI-101 AN-GUS and pRI-101 AN-GUS-target genes (*NAC*, *APE*, *4CL*, and *PAL*). The GUS fluorescence values of tobacco leaves co-transformed by pRI-101 AN-miRNA (miR164d, miR396b, Novel-miR1, and Novel-miR7) and pRI-101 AN-GUS-target gene (*NAC*, *APE*, *4CL*, and *PAL*) were determined. The quantitative GUS fluorescence analysis supported the results of GUS histochemical staining. These experiments indicated the existence of a negative regulatory relationship between candidate miRNAs and target genes.

Transient Expression Levels of Candidate miRNAs and Target Genes

In this experiment, miR164d, miR396b, Novel-miR1, Novel-miR7, Novel-miR1/Novel-miR7 (1:1, v/v), *NAC*, *APE*, *4CL*, *PAL*, and *4CL/PAL* (1:1, v/v) were transiently expressed in cucumber cotyledons via *Agrobacterium* infiltration. Normal tobacco (Control) and tobacco injected with pRI-101 AN were used as controls. Expression levels of candidate miRNAs and target genes in transgenic cucumbers were detected by qRT-PCR. The results are shown in **Figure 3**. The gene expression levels were similar in the normal tobacco and pRI-101 AN experimental groups. The gene expression level of the overexpression group was significantly higher than that of the controls, which proved that the transient transformation was successful and could be used for further experiments.

Function of Candidate miRNAs and Target Genes in the Interaction Between Cucumber and *C. cassiicola*

We performed analyses of transient expression in cucumber cotyledons to explore the functions of candidate miRNAs
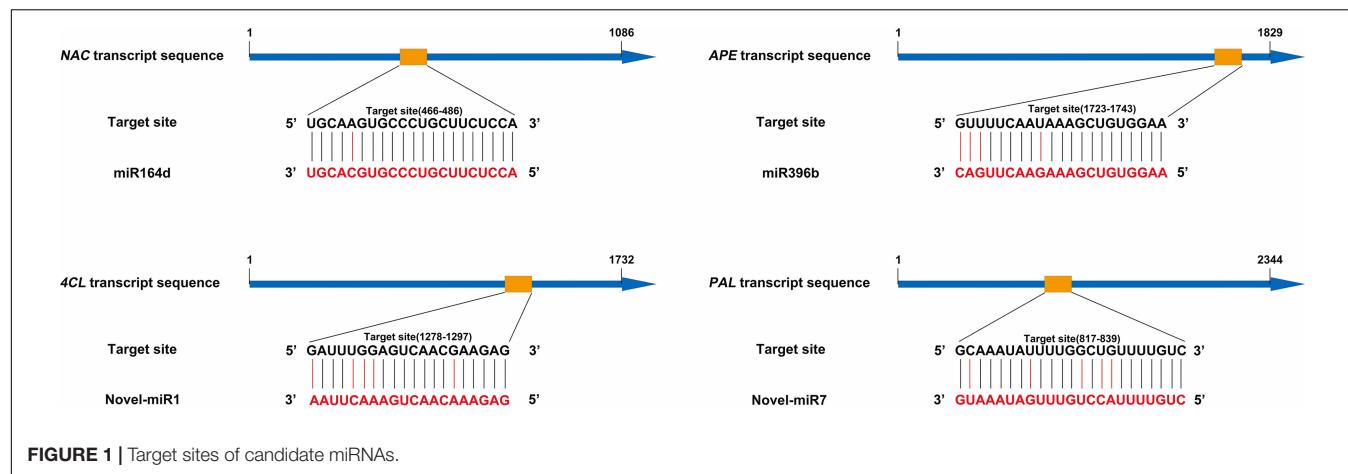
FIGURE 1 | Target sites of candidate miRNAs.
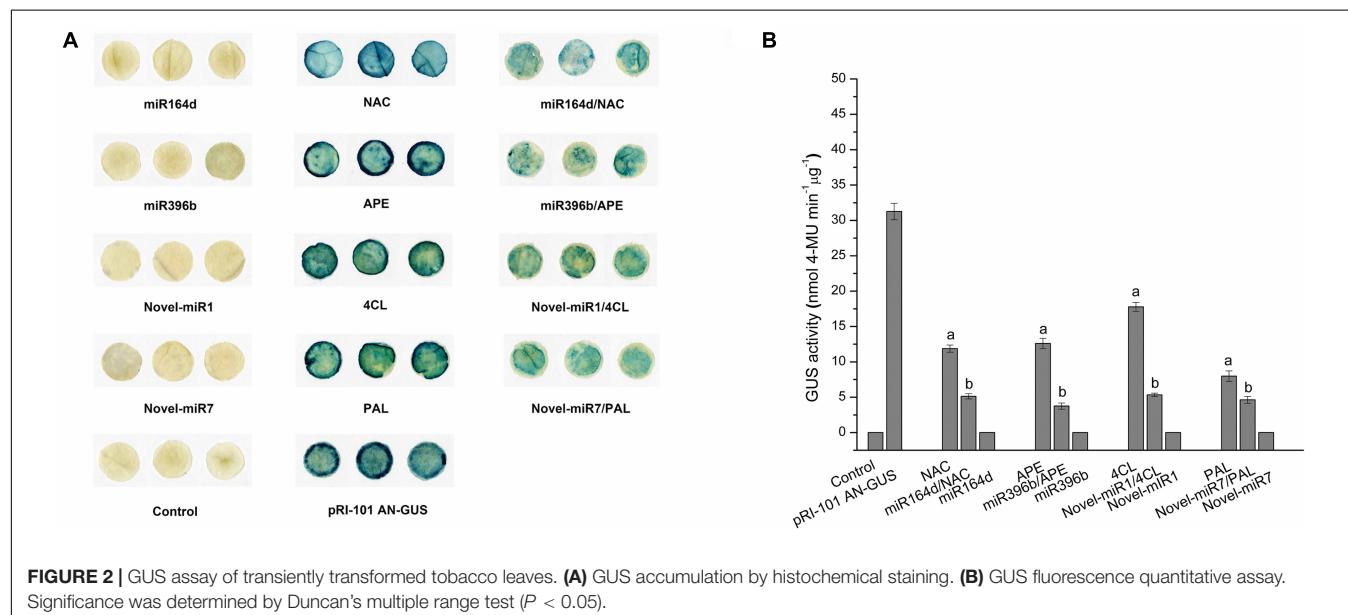
FIGURE 2 | GUS assay of transiently transformed tobacco leaves. **(A)** GUS accumulation by histochemical staining. **(B)** GUS fluorescence quantitative assay. Significance was determined by Duncan's multiple range test ($P < 0.05$).

and target genes in disease resistance to *C. cassiicola*. Normal cucumber cotyledons (Control) and cucumber cotyledons injected with pRI-101 AN were used as controls. Two days after agroinfiltration for transient expression, the agroinfiltrated cotyledons were collected for disease assays by dropping spore suspensions of *C. cassiicola* onto cucumber cotyledons. The disease phenotype at 5 days after inoculation showed that the *C. cassiicola*-induced lesions in the experimental groups of miR164d, miR396b, Novel-miR7, and Novel-miR1/Novel-miR7 were significantly larger than those in the controls (**Figures 4A,B**), indicating that transient expression of these miRNAs in cucumber cotyledons reduced the resistance to *C. cassiicola*. However, the *C. cassiicola*-induced lesions on *NAC*-, *APE*-, *4CL*-, and *4CL/PAL*-infiltrated leaves were markedly smaller (**Figures 4A,B**) than those of the controls, showing that transient expression of these genes in cucumber cotyledons improved the resistance to *C. cassiicola*. Expression of the *C. cassiicola Actin* gene

was used as a standard of fungal growth. The growth of *C. cassiicola* in the miR164d-, miR396b-, Novel-miR1-, and Novel-miR7-infiltrated leaves was significantly higher than that of the control, while the growth in the *NAC*-, *APE*-, *4CL*-, and *4CL/PAL*-infiltrated leaves was markedly lower than that of the control (**Figure 4C**). In addition, the disease lesions of the Novel-miR1 group did not increase significantly, but the growth of *C. cassiicola* increased significantly (**Figures 4A–C**). Transient expression assays showed that overexpression of candidate miRNAs could reduce the resistance to *C. cassiicola*, and overexpression of target genes corresponding to candidate miRNAs could improve the resistance to *C. cassiicola*. These results were consistent with those of our previous analysis.

The previous analysis shows that Novel-miR1 and Novel-miR7 can inhibit *4CL* and *PAL*, respectively. The *4CL* and *PAL* genes are two key genes in the lignin synthesis pathway, and the lignin content is positively correlated with disease
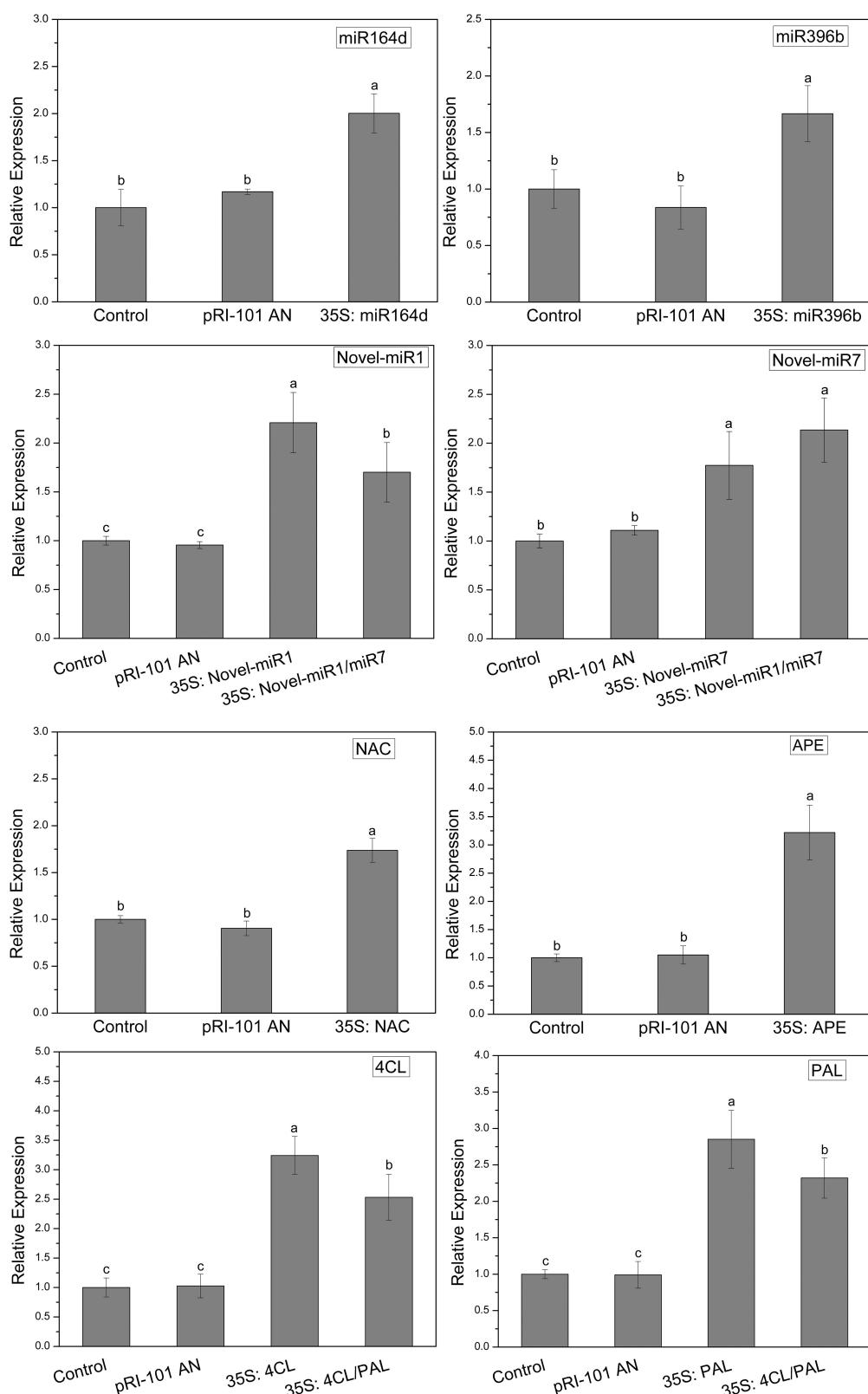
FIGURE 3 | Transient expression levels of miRNAs and target genes in cucumber cotyledons. Significance was determined by Duncan's multiple range test ($P < 0.05$).

FIGURE 4 | Disease phenotype in cucumber cotyledons transiently expressing miRNAs and target genes after inoculation with *C. cassiicola*. **(A)** Disease symptoms of cucumber cotyledons. **(B)** Disease lesion size of cucumber cotyledons. **(C)** Biomass of *C. cassiicola* of cucumber cotyledons. **(D)** Effect of transient expression of miRNAs and targets on lignin content in cucumber cotyledons. Significance was determined by Duncan's multiple range test ($P < 0.05$).
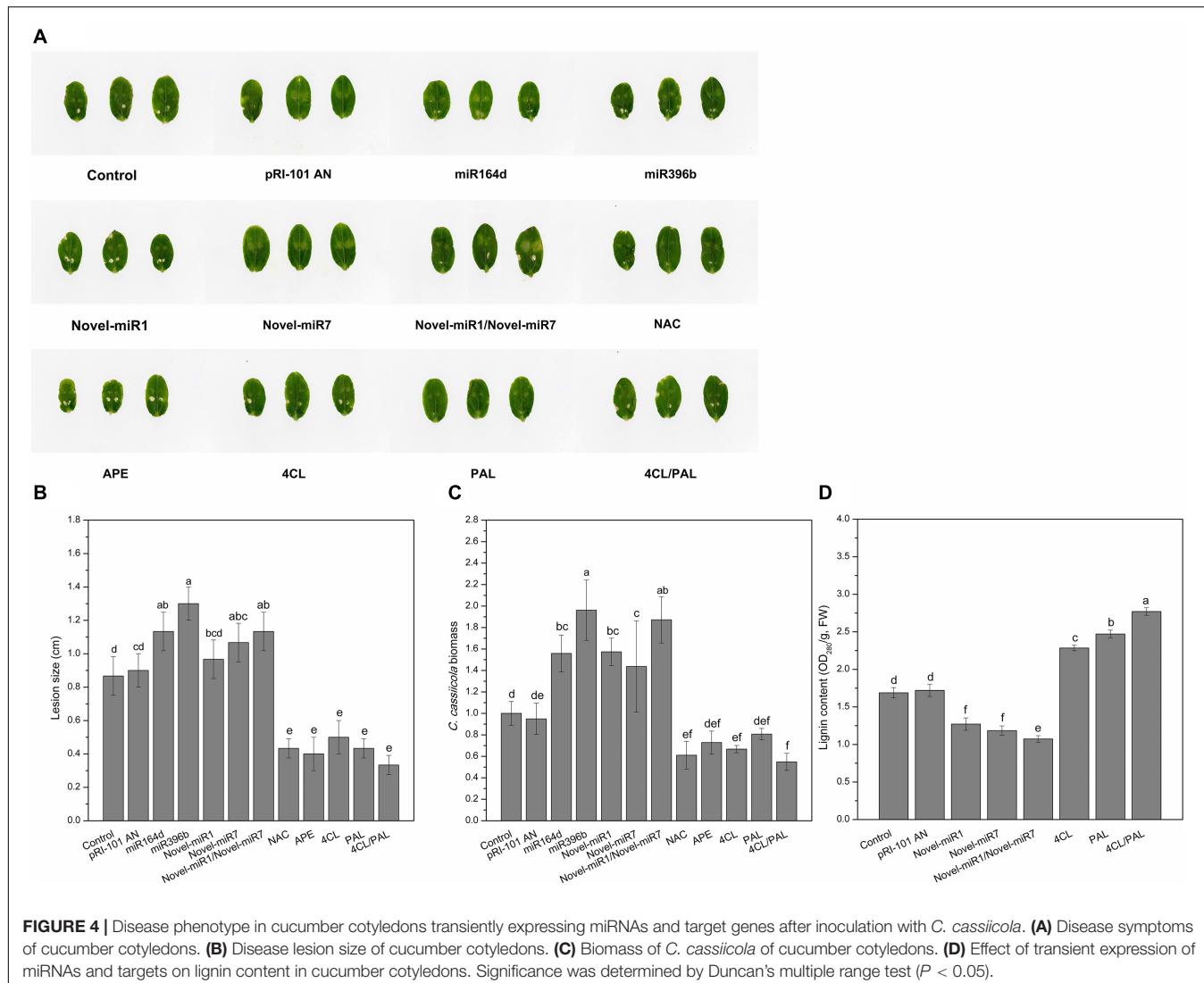
resistance. In this experiment, Novel-miR1, Novel-miR7, Novel-miR1/Novel-miR7 (1:1, v/v), *4CL*, *PAL*, and *4CL/PAL* (1:1, v/v) were transiently expressed in cucumber cotyledons through agroinfiltration. Normal tobacco (Control) and tobacco injected with pRI-101 AN were used as controls. Two days after agroinfiltration for transient expression, the agroinfiltrated cotyledons were prepared for determination of lignin content (**Figure 4D**). The lignin content was determined to further analyze the effect of overexpression of candidate miRNAs and target genes on lignin accumulation in cucumber cotyledons. As shown in **Figure 4D**, the lignin content was significantly increased in *4CL*-, *PAL*-, and *4CL/PAL*-infiltrated samples compared to that of the control, and the lignin content in Novel-miR1-, Novel-miR7-, and Novel-miR1/Novel-miR7-infiltrated leaves was obviously decreased compared to that of the control. The lignin content of *4CL/PAL*-infiltrated leaves was the highest, while the lignin content of Novel-miR1/Novel-miR7-infiltrated leaves was the lowest. The results showed that overexpression of *4CL* and *PAL* could increase

the lignin content in cucumber leaves, and overexpression of Novel-miR1 and Novel-miR7 could reduce lignin in cucumber leaves.

TRV-Induced VBMS

In this study, STTM-miR164d, STTM-miR396b, STTM-Novel-miR1, and STTM-Novel-miR7 recombinant vectors were constructed based on TRV. The STTM structures are shown in **Figure 5A**. The sequences of STTM-miR164d, STTM-miR396b, STTM-Novel-miR1, and STTM-Novel-miR7 used in this study are listed in **Supplementary Table S1**. The sides of the structure are TMs, and the restriction sites at both ends are *Eco*RI and *Sac*I. STTM and pTRV2 were used to construct recombinant vectors by In-Fusion technology. Recombinant viruses TRV: 00 (pTRV1+pTRV2), TRV: STTM-miR164d, TRV: STTM-miR396b, TRV: STTM-Novel-miR1, and TRV: STTM-Novel-miR7 were infiltrated into cucumber cotyledons. After inoculation for 7 days, chlorosis and a few virus spots appeared in the cotyledons of cucumber inoculated with TRV recombinant
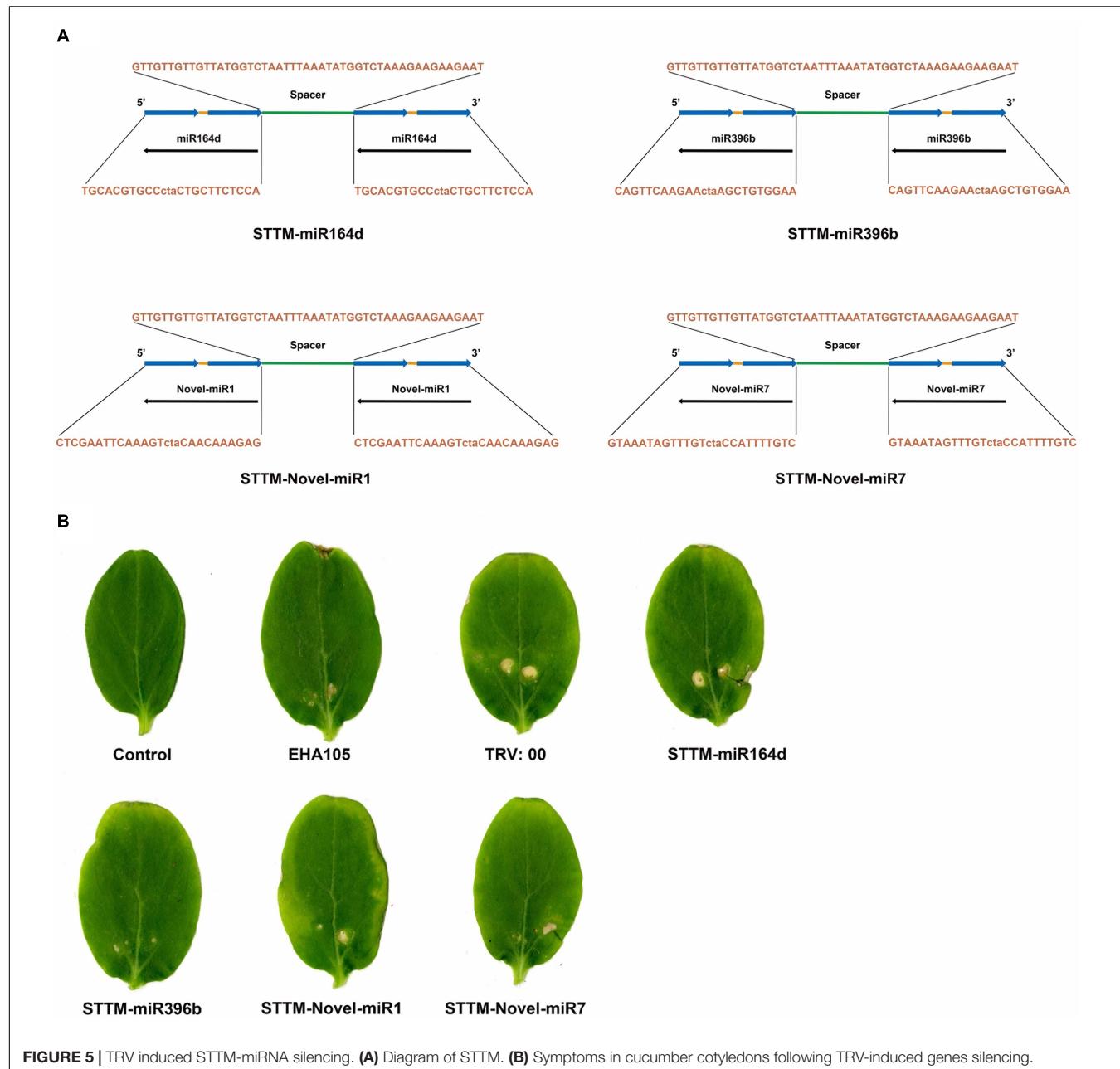
FIGURE 5 | TRV induced STTM-miRNA silencing. **(A)** Diagram of STTM. **(B)** Symptoms in cucumber cotyledons following TRV-induced genes silencing.

virus, but no obvious phenotype was observed on the cotyledons of Control and *A. tumefaciens* EHA105 (**Figure 5B**), indicating that TRV had successfully replicated and proliferated in the cotyledons of cucumber.

To explore the silencing efficacy of candidate miRNAs, we detected the expression levels of candidate miRNAs and corresponding target genes in cucumber cotyledons injected with TRV: STTM by qRT-PCR. As shown in **Figure 6**, the expression levels of candidate miRNAs in the cotyledons infiltrated with TRV: STTM-miR164d, TRV: STTM-miR396b, TRV: STTM-Novel-miR1, and TRV: STTM-Novel-miR7 were significantly decreased, and the expression levels of target genes corresponding to these miRNAs were upregulated. These results

indicated that candidate miRNAs in cucumber cotyledons were successfully inhibited.

Response of Cucumber Cotyledons to *C. cassiicola* After Candidate miRNAs Silencing

To identify the role of STTM-miRNA in the interaction between cucumber and *C. cassiicola*, we inoculated the TRV: 00 (pTRV1+pTRV2)-, TRV: STTM-miR164d-, TRV: STTM-miR396b-, TRV: STTM-Novel-miR1-, and TRV: STTM-Novel-miR7-infiltrated cucumber cotyledons with *C. cassiicola* to observe phenotypic changes. After 5 days of inoculation
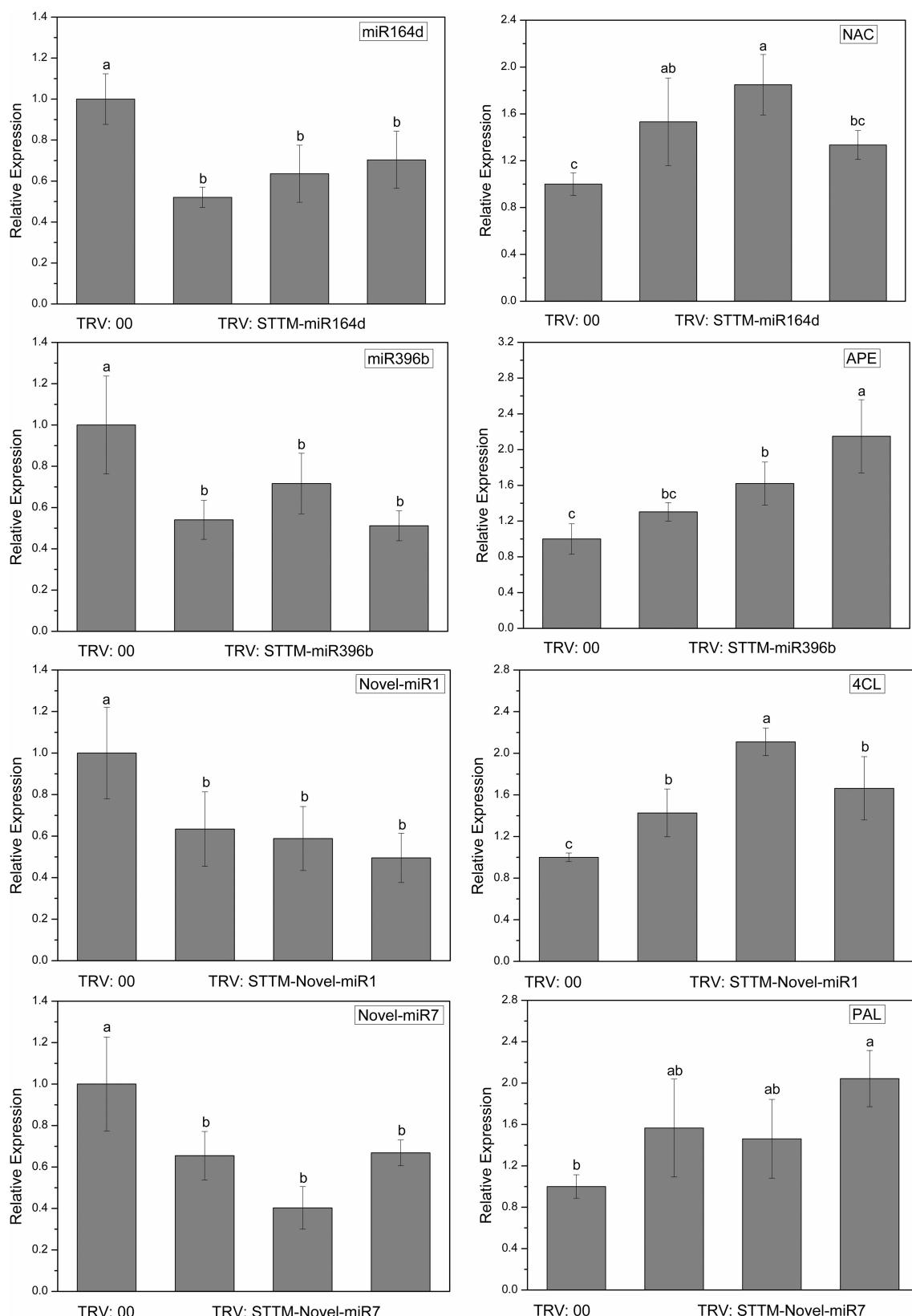
FIGURE 6 | Expression levels of miRNAs and targets in cucumber cotyledons. Significance was determined by Duncan's multiple range test ($P < 0.05$).

FIGURE 7 | Disease phenotype in STTM transgenic cucumber cotyledons after inoculation with *C. cassiicola*. **(A)** Disease symptoms of cucumber cotyledons. **(B)** Disease lesion size of cucumber cotyledons. **(C)** Biomass of *C. cassiicola* of cucumber cotyledons. **(D)** Effect of candidate miRNAs silencing on lignin content in cucumber cotyledons. Significance was determined by Duncan's multiple range test ($P < 0.05$).
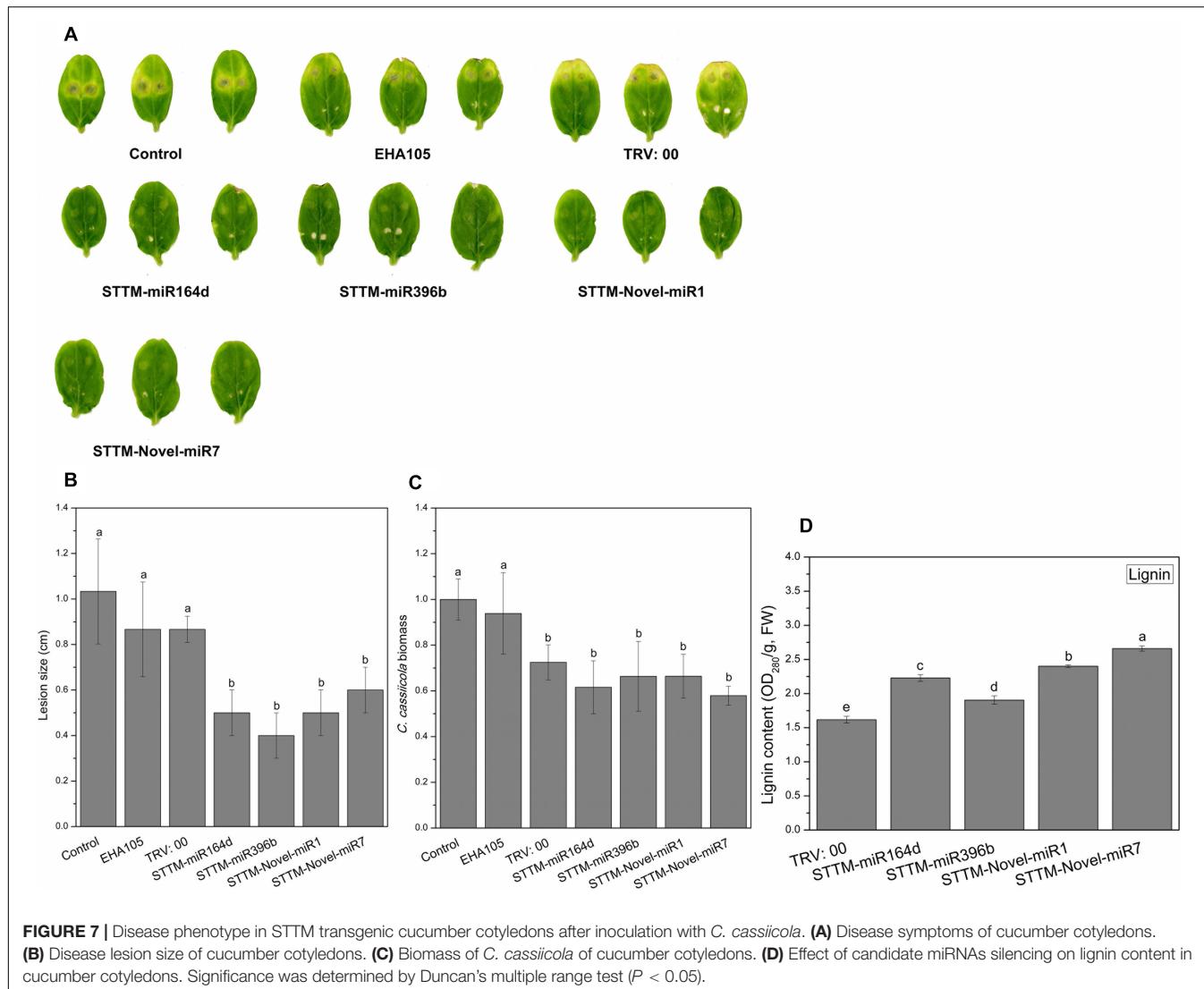
of *C. cassiicola*, there was no significant difference between the disease lesions of TRV: 00 (pTRV1+pTRV2)-infiltrated cotyledons and control cotyledons. However, the *C. cassiicola*-induced lesions of miR164d-, miR396b-, Novel-miR1-, and Novel-miR7-silenced cucumber cotyledons were markedly smaller (**Figures 7A,B**) than those of the control. The biomass of *C. cassiicola* in the infected cotyledons of cucumber was detected by qRT-PCR (**Figure 7C**), and the expression of the *C. cassiicola Actin* gene in candidate miRNA-silenced cucumber cotyledons decreased significantly. Thus, silencing of miR164d, miR396b, Novel-miR1 and Novel-miR7 increased cucumber resistance to *C. cassiicola*. The lignin content was determined to further analyze the effect of silencing candidate miRNAs on lignin accumulation in cucumber cotyledons (**Figure 7D**). The lignin content was significantly upregulated in all silenced plants, especially in Novel-miR1- and Novel-miR7-silenced plants. Because *4CL* (inhibited by Novel-miR1) and *PAL* (inhibited by Novel-miR7) are upstream and downstream genes in the phenylpropane pathway, which is related to lignin

synthesis, this finding is also consistent with our previous experimental results.

DISCUSSION

As miRNAs have no coding function, they can only act by inhibiting or degrading the corresponding targets. Genes related to abiotic and biotic stresses have been proven to be targets of miRNAs and can be used to determine the function of miRNAs (Jagadeeswaran et al., 2009; Katiyar-Agarwal and Jin, 2010). In plants, the interaction between miRNAs and target genes was verified by 5′ RNA ligase-mediated rapid amplification of cDNA ends (5′ RLM-RACE) (Llave et al., 2002). The RACE method is cumbersome and does not visualize the interaction between the miRNA and target gene. *Agrobacterium*-mediated transient expression of tobacco is highly efficient and has a long expression time, which is suitable for the study of gene interactions in plants (Prabu and Prasad, 2012; Yin et al., 2017). This method is mainly
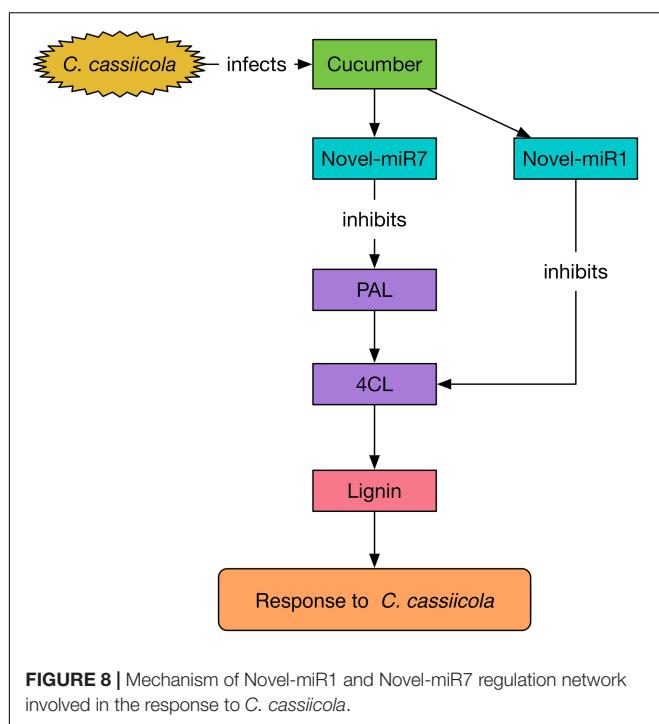
FIGURE 8 | Mechanism of Novel-miR1 and Novel-miR7 regulation network involved in the response to *C. cassiicola*.

used to verify the interaction between miRNAs and target genes by detecting the expression of the *GUS* gene in tobacco after transient expression of miRNAs and target genes (Feng et al., 2013; Han et al., 2016).

NAC is a plant-specific transcription factor that is involved in plant development and stress responses in plants. Previous studies have shown that many members of the *NAC* gene family exhibit differential expression characteristics after pathogen infection in plants, which indicates that *NAC* genes have specific biological functions in response to disease resistance in plants (Jensen and Skriver, 2014; Qin et al., 2014). In a recent study, researchers analyzed cucumber *NAC* genes related to the response to abiotic stresses (Zhang et al., 2017). However, there are few studies on the role of *NAC* genes in the interaction between cucumber and pathogens. BLAST analysis of existing data showed that the *NAC* gene we identified was *csNAC30*. Some studies have shown that *NAC* affects the synthesis of secondary metabolites, but the detailed mechanism is not clear (Zhao et al., 2010; Xu et al., 2015).

Tryptophan not only promotes endogenous jasmonic acid (JA) biosynthesis and triggers plant signal transduction pathways but also participates in the terpenoid indole alkaloid (TIA) pathway, which in turn regulates the plant response to stress (Sun et al., 2016). Anthranilate phosphoribosyltransferase (APE) is an important enzyme in the tryptophan synthesis pathway, which is indirectly involved in the synthesis of plant secondary metabolites (Dharmawardhana et al., 2013).

Phenylalanine ammonia-lyase (*PAL*) and *4-coumarate: CoA ligase* (*4CL*) are two key genes in the phenylpropane synthesis pathway and are closely related to plant resistance to external stresses. First, PAL catalyzes phenylalanine to cinnamic acid;

then, cinnamic acid participates in the synthesis of ubiquinone and is catalyzed by 4CL to form cinnamoyl-CoA; finally, cinnamoyl-CoA further synthesizes lignin and flavonoids (Kim and Hwang, 2014; Li et al., 2015).

After analyzing the function of the target genes, we selected miR164d, miR396b, Novel-miR1, and Novel-miR7 and their target genes *NAC*, *APE*, *4CL*, and *PAL* for interaction verification. Histochemical staining and quantitative detection of *GUS* genes showed that these target genes could be suppressed by the corresponding miRNA, but the inhibition efficiency was not 100%. There are two possible causes for this result. First, there is a balance between the transcription of target genes and nucleotide cleavage, and this balance is affected by the expression level of miRNAs (Nikovics et al., 2006). Second, it may be that only a portion of the target mRNA was cleaved and degraded by the miRNA, and the remaining target mRNAs were detached from the cleavage system and normally transcribed (Adam et al., 2010).

To explore the functions of miR164d, miR396b, Novel-miR1, and Novel-miR7 and their target genes *NAC*, *APE*, *4CL*, and *PAL* in cucumber response to *C. cassiicola*, we transiently expressed candidate miRNAs and target genes in cucumber cotyledons. After vaccination with *C. cassiicola*, phenotypic changes were observed. Because *4CL* and *PAL* are upstream and downstream genes in the phenylpropanoid metabolic pathway, we coexpressed *4CL* and *PAL* to establish the experimental group. Similarly, Novel-miR1 and Novel-miR7 were also coexpressed. **Figure 4** shows that the disease resistance of the target gene overexpression groups was significantly higher than that of the control groups, indicating that these genes play an important role in cucumber resistance to *C. cassiicola*. The disease resistance of the candidate miRNAs overexpression group was lower than that of the control group, probably because these miRNAs inhibited the expression of their target genes. These results were also confirmed by measurement of the biomass of *C. cassiicola* in cucumber cotyledons after infection. The data showed that the coexpressed *4CL* and *PAL* groups had the highest disease resistance, and the disease resistance of the corresponding Novel-miR1 and Novel-miR7 coexpression groups was the lowest. Because the phenylpropanoid metabolic pathway affects the synthesis of lignin, and *4CL* and *PAL* are important genes in this pathway, we performed lignin content determination after transient expression of candidate miRNAs and genes in cucumber cotyledons for 2 days. **Figure 4D** shows that the lignin content of *4CL/PAL*-infiltrated leaves was the highest; the lignin content of Novel-miR1-, Novel-miR7-, and Novel-miR1/Novel-miR7-infiltrated leaves was lower than that of the normal tobacco, and the lignin content reduction was more severe in the Novel-miR1/Novel-miR7-infiltrated leaves. STTM is an effective tool for studying miRNA function in plants and animals. Tang et al. (2012) developed STTM technology based TM. Overexpression of STTM has been used to identify the functions of miRNAs in a variety of plants, such as *Arabidopsis* and wheat (Jia et al., 2015; Jiao et al., 2015), but it has not been reported in cucumber. To further explore the function of candidate miRNAs, we developed a TRV-induced VBMS to silence cucumber endogenous miRNAs. The results showed that silencing of miR164d, miR396b, Novel-miR1 and Novel-miR7 increased the resistance of cucumber

cotyledons to *C. cassiicola*, indicating that these miRNAs played a negative regulatory role in cucumber resistance to *C. cassiicola*. Notably, lignin content is highest in Novel-miR1- and Novel-miR7-silenced cucumber cotyledons. These findings also indicate that the interaction between Novel-miR1 and Novel-miR7 and their target genes affects the synthesis of lignin, which in turn affects the resistance of cucumber to *C. cassiicola* (**Figure 8**).

CONCLUSION

In our study, the expression vectors were constructed by In-Fusion technology, and the negative relationships of miR164d, miR396b, Novel-miR1, and Novel-miR7 and their targets were verified by the *Agrobacterium*-mediated tobacco transient expression system. Meanwhile, we found that overexpression of *NAC*, *APE*, *4CL*, and *PAL* could improve the resistance to *C. cassiicola*, and overexpression of miR164d, miR396b, Novel-miR1, and Novel-miR7 could reduce the disease resistance in cucumber. We confirmed that silencing candidate miRNAs could improve the disease resistance of cucumber, and the lignin content in Novel-miR1- and Novel-miR7-silenced cucumber cotyledons was significantly increased. These candidate miRNAs and targets are closely related to cucumber lignin synthesis, and the data combined with previous analyses demonstrate the important role of secondary metabolism, especially the lignin metabolism pathway, in the process of cucumber resistance to *C. cassiicola*.

DATA AVAILABILITY

AUTHOR CONTRIBUTIONS

XW and HF conceived and designed the research. XW performed the experiments and wrote the manuscript. All authors analyzed the data and read and approved the final manuscript. HF revised the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00668/full#supplementary-material

REFERENCES

Adam, H., Marguerettaz, M., Qadri, R., Adroher, B., Richard, F., Collin, M., et al. (2010). Divergent expression patterns of *miR164* and *CUP-SHAPED COTYLEDON* genes in palms and other monocots: implication for the evolution of meristem function in angiosperms. *Mol. Biol. Evol.* 28, 1439–1454. doi: 10.1093/molbev/msq328

Agrawal, A. A., and Weber, M. G. (2015). On the study of plant defence and herbivory using comparative approaches: how important are secondary plant compounds. *Ecol. Lett.* 18, 985–991. doi: 10.1111/ele.12482

Bachan, S., and Dinesh-Kumar, S. P. (2012). Tobacco rattle virus (TRV)-based virus-induced gene silencing. *Methods Mol. Biol.* 89, 83–92. doi: 10.1007/978-1-61779-882-5_6

Baxter, A., Mittler, R., and Suzuki, N. (2014). ROS as key players in plant stress signalling. *J. Exp. Bot.* 65, 1229–1240. doi: 10.1093/jxb/ert375

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Burkhardt, A., and Day, B. (2016). Transcriptome and small RNAome dynamics during a resistant and susceptible interaction between cucumber and downy mildew. *Plant Genome* 9, 1–19. doi: 10.3835/plantgenome2015.08.0069

Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D., et al. (2013). A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice* 6:15. doi: 10.1186/1939-8433-6-15

Feng, H., Zhang, Q., Wang, Q., Wang, X., Liu, J., Li, M., et al. (2013). Target of tae-miR408, a chemocyanin-like protein gene (*TaCLP1*), plays positive roles in wheat response to high-salinity, heavy cupric stress and stripe rust. *Plant Mol. Biol.* 83, 433–443. doi: 10.1007/s11103-013-0101-9

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079

Han, L., Weng, K., Ma, H., Xiang, G., Li, Z., Wang, Y., et al. (2016). Identification and characterization of *erysiphe necator*-responsive microRNAs in Chinese wild *vitis pseudoreticulata* by high-throughput sequencing. *Front. Plant Sci.* 7:621. doi: 10.3389/fpls.2016.00621

Huang, C. J., Qian, Y. J., Li, Z. H., and Zhou, X. P. (2012). Virus-induced gene silencing and its application in plant functional genomics. *Sci. China Life Sci.* 55, 99–108. doi: 10.1007/s11427-012-4280-4

Ishii, H., Yano, K., Date, H., Furuta, A., Sagehashi, Y., Yamaguchi, T., et al. (2007). Molecular characterization and diagnosis of QoI resistance in cucumber and eggplant fungal pathogens. *Phytopathology* 97, 1458–1466. doi: 10.1094/PHYTO-97-11-1458

Jagadeeswaran, G., Saini, A., and Sunkar, R. (2009). Biotic and abiotic stress down-regulate miR398 expression in *Arabidopsis*. *Planta* 229, 1009–1014. doi: 10.1007/s00425-009-0889-3

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6, 3901–3907. doi: 10.1089/dna.1987.6.583

Jensen, M. K., and Skriver, K. (2014). NAC transcription factor gene regulatory and protein-protein interaction networks in plant stress responses and senescence. *IUBMB Life* 66, 156–166. doi: 10.1002/iub.1256

Jia, X. Y., Ding, N., Fan, W. X., Yan, J., Gu, Y. Y., Tang, X. Q., et al. (2015). Functional plasticity of miR165/166 in plant development revealed by small tandem target mimic. *Plant Sci.* 233, 11–21. doi: 10.1016/j.plantsci.2014.12.020

Jiang, N., Meng, J., Cui, J., Sun, G., and Luan, Y. (2018). Function identification of miR482b, a negative regulator during tomato resistance to *Phytophthora infestans*. *Hortic. Res.* 5:9. doi: 10.1038/s41438-018-0017-2

Jiao, J., Wang, Y., Selvaraj, J. N., Xing, F., and Liu, Y. (2015). Barley stripe mosaic virus (BSMV) induced microRNA silencing in common wheat (*Triticum aestivum* L.). *PLoS One* 10:e0126621. doi: 10.1371/journal.pone.0126621

Jin, W. B., and Wu, F. L. (2015). Identification and characterization of cucumber microRNAs in response to *Pseudoperonospora cubensis* infection. *Gene* 569, 225–232. doi: 10.1016/j.gene.2015.05.064

Katiyar-Agarwal, S., and Jin, H. (2010). Role of small RNAs in host-microbe interactions. *Annu. Rev. Phytopathol.* 48, 225–246. doi: 10.1146/annurev-phyto-073009-114457

Khraiwesh, B., Zhu, J. K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *BBA Gene Regul. Mech.* 1819, 137–148. doi: 10.1016/j.bbagrm.2011.05.001

Kim, D. S., and Hwang, B. K. (2014). An important role of the pepper phenylalanine ammonia-lyase gene (*PAL1*) in salicylic acid-dependent signalling of the defence response to microbial pathogens. *J. Exp. Bot.* 65, 2295–2306. doi: 10.1093/jxb/eru109

Kwon, M. K., Kang, B. R., Cho, B. H., and Kim, Y. C. (2003). Occurrence of target leaf spot disease caused by *Corynespora cassicolaon* cucumber in Korea. *Plant Pathol.* 52:424. doi: 10.1046/j.1365-3059.2003.00846.x

Li, B., Gao, W., Shi, Y. X., and Xie, X. W. (2012). Progress in researches on *Corynespora* leaf spot. *Acta Phytophylacica Sin.* 39, 171–176.

Li, C. H., Li, Y., Bai, L., Zhang, T., He, C., Yan, Y., et al. (2014). Grafting-responsive miRNAs in cucumber and pumpkin seedlings identified by high-throughput sequencing at whole genome level. *Physiol. Plant.* 151, 406–422. doi: 10.1111/ppl.12122

Li, Y., Kim, J. I., Pysh, L., and Chapple, C. (2015). Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol.* 169, 2409–2421. doi: 10.1104/pp.15.00838

Liu, Q., Wang, H., Hu, H., and Zhang, H. (2015). Genome-wide identification and evolutionary analysis of positively selected miRNA genes in domesticated rice. *Mol. Genet. Genom.* 290, 593–602. doi: 10.1007/s00438-014-0943-0

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

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 297, 2053–2056. doi: 10.1126/science.1076311

MacFarlane, S. A., Vassilakos, N., and Brown, D. J. (1999). Similarities in the genome organization of tobacco rattle virus and pea early-browning virus isolates that are transmitted by the same vector nematode. *J. Gen. Virol.* 80, 273–276. doi: 10.1099/0022-1317-80-1-273

Mao, W. H., Li, Z. Y., Xia, X. J., Li, Y. D., and Yu, J. Q. (2012). A combined approach of high-throughput sequencing and degradome analysis reveals tissue specific expression of microRNAs and their targets in cucumber. *PLoS One* 7:e33040. doi: 10.1371/journal.pone.0033040

Martínez, G., Forment, J., Llave, C., Pallás, V., and Gómez, G. (2011). High-throughput sequencing, characterization and detection of new and conserved cucumber miRNAs. *PLoS One* 6:e19523. doi: 10.1371/journal.pone.0019523

Meng, Y., Shao, C., Wang, H., and Jin, Y. (2012). Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. *BMC Genom.* 13:197. doi: 10.1186/1471-2164-13-197

Miyamoto, T., Ishii, H., Stammler, G., Koch, A., Ogawara, T., Tomita, Y., et al. (2010). Distribution and molecular characterization of *Corynespora cassiicola* isolates resistant to boscalid. *Plant Pathol.* 59, 873–881. doi: 10.1111/j.1365-3059.2010.02321.x

Morrison, I. M. (1972). A semi-micro method for the determination of lignin and its use in predicting the digestibility of forage crops. *J. Sci. Food Agr.* 23, 455–463. doi: 10.1002/jsfa.2740230405

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 312, 436–439. doi: 10.1126/science.1126088

Nghia, N. A., Kadir, J., Sunderasan, E., Puad, A. M., Malik, A., and Napis, S. (2008). Morphological and inter simple sequence repeat (ISSR) markers analyses of *Corynespora cassiicola* isolates from rubber plantations in Malaysia. *Mycopathologia* 166, 189–201. doi: 10.1007/s11046-008-9138-8

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., et al. (2006). The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* 18, 2929–2945. doi: 10.1105/tpc.106.045617

Pateraki, I., and Kanellis, A. K. (2010). Stress and developmental responses of terpenoid biosynthetic genes in *Cistus creticus* subsp. *creticus*. *Plant Cell Rep.* 29, 629–641. doi: 10.1007/s00299-010-0849-1

Prabu, G., and Prasad, D. T. (2012). Functional characterization of sugarcane MYB transcription factor gene promoter (*PScMYBAS1*) in response to abiotic stresses and hormone. *Plant Cell Rep.* 31, 661–669. doi: 10.1007/s00299-011-1183-y

Pusztahelyi, T., Holb, I. J., and Pócsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Front. Plant Sci.* 6:573. doi: 10.3389/fpls.2015.00573

Qi, Y. X., Zhang, X., Pu, J. J., Liu, X. M., Lu, Y., Zhang, H., et al. (2011). Morphological and molecular analysis of genetic variability within isolates of *Corynespora cassiicola* from different hosts. *Eur. J. Plant Pathol.* 130, 83–95. doi: 10.1007/s10658-010-9734-6

Qin, X., Zheng, X., Huang, X., Lii, Y., Shao, C., Xu, Y., et al. (2014). A novel transcription factor JcNAC1 response to stress in new model woody plant *Jatropha curcas*. *Planta* 239, 511–520. doi: 10.1007/s00425-013-1993-y

Sanan-Mishra, N., Kumar, V., Sopory, S. K., and Mukherjee, S. K. (2009). Cloning and validation of novel miRNA from basmati rice indicates cross talk between abiotic and biotic stresses. *Mol. Genet. Genom.* 282, 463. doi: 10.1007/s00438-009-0478-y

Sha, A., Zhao, J., Yin, K., Tang, Y., Wang, Y., Wei, X., et al. (2014). Virus-based microRNA silencing in plants. *Plant Physiol.* 164, 36–47. doi: 10.1104/pp.113.23110

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans heterochronic* pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* 5, 659–669. doi: 10.1016/S1097-2765(00)80245-2

Sun, J. Y., Manmathan, H., Sun, C., and Peebles, C. A. M. (2016). Examining the transcriptional response of overexpressing anthranilate synthase in the hairy roots of an important medicinal plant *Catharanthus roseus* by RNA-seq. *BMC Plant Biol.* 16:108. doi: 10.1186/s12870-016-0794-4

Tang, G. L., Yan, J., Gu, Y. Y., Qiao, M. M., Fan, R. W., Mao, Y. P., et al. (2012). Construction of short tandem target mimic (STTM) to block the functions of plant and animal microRNAs. *Methods* 58, 118–125. doi: 10.1016/j.ymeth.2012.10.006

Teramoto, A., Martins, M. C., Ferreira, L. C., and Cunha, M. G. (2011). Reaction of hybrids, inhibition in vitro and target spot control in cucumber. *Hortic. Bras.* 29, 342–348. doi: 10.1590/S0102-05362011000300014

Thiebaut, F., Grativol, C., Hemerly, A. S., and Ferreira, P. C. G. (2015). MicroRNA networks in plant-microorganism interactions. *Trop. Plant Biol.* 8, 40–50. doi: 10.1007/s12042-015-9149-9

Wang, X., Zhang, D., Cui, N., Yu, Y., Yu, G., and Fan, H. (2018). Transcriptome and miRNA analyses of the response to *Corynespora cassiicola* in cucumber. *Sci. Rep.* 8:7798. doi: 10.1038/s41598-018-26080-6

Waterhouse, P. M., and Hellens, R. P. (2015). Coding in non-coding RNAs. *Nature* 520, 41–42. doi: 10.1038/nrg3074

Wen, C., Mao, A., Dong, C., Liu, H., Yu, S., Guo, Y. D., et al. (2015). Fine genetic mapping of target leaf spot resistance gene *cca-3* in cucumber, *Cucumis sativus* L. *Theor. Appl. Genet.* 128, 2495–2506. doi: 10.1007/s00122-015-2604-z

Xu, Q., Wang, W., Zeng, J., Zhang, J., Grierson, D., Li, X., et al. (2015). A NAC transcription factor, *EjNAC1*, affects lignification of loquat fruit by regulating lignin. *Postharvest Biol. Technol.* 102, 25–31. doi: 10.1016/j.postharvbio.2015.02.002

Yan, J., Gu, Y., Jia, X., Kang, W., Pan, S., Tang, X., et al. (2012). Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* 24, 415–427. doi: 10.1105/tpc.111.094144

Yang, L. M., Koo, D. H., Li, Y. H., Zhang, X. J., Luan, F. S., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yin, J. L., Wong, W. S., Jang, I. C., and Chua, N. H. (2017). Co-expression of peppermint geranyl diphosphate synthase small subunit enhances monoterpene production in transgenic tobacco plants. *New Phytol.* 213, 1133–1144. doi: 10.1111/nph.14280

Zhang, B., and Wang, Q. (2015). MicroRNA-based biotechnology for plant improvement. *J. Cell Physiol.* 230, 1–15. doi: 10.1002/jcp.24685

Zhang, X., Zhao, H., Gao, S., Wang, W. C., Katiyar-Agarwal, S., Huang, H. D., et al. (2011). *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393*-mediated silencing of a Golgi-localized snare gene, MEMB12. *Mol. Cell* 42, 356–366. doi: 10.1016/j.molcel.2011.04.010

Zhang, X. M., Yu, H. J., Sun, C., Deng, J., Zhang, X., Liu, P., et al. (2017). Genome-wide characterization and expression profiling of the *NAC* genes under abiotic stresses in *Cucumis sativus*. *Plant Physiol. Biochem.* 113, 98–109. doi: 10.1016/j.plaphy.2017.01.023

Zhao, Q., and Dixon, R. A. (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 16, 227–233. doi: 10.1016/j.tplants.2010.12.005

Zhao, Q., Gallego-Giraldo, L., Wang, H., Zeng, Y., Ding, S. Y., Chen, F., et al. (2010). An NAC transcription factor orchestrates multiple features of cell wall development in *Medicago truncatula*. *Plant J.* 63, 100–114. doi: 10.1111/j.1365-313X.2010.04223.x

Zhou, X. F., Sun, J. D., Zhao, Z., Lv, J., Wei, X. W., Cai, R., et al. (2012). The feasibility analysis of PVX and TRV vectors as the VIGS tool for studying the gene function. *Phys. Procedia* 33, 46–54. doi: 10.1016/j.phpro.2012.05.029

Chromosomal Locations and Interactions of Four Loci Associated With Seed Coat Color in Watermelon

Lucky Paudel[1], Josh Clevenger[1] and Cecilia McGregor[1,2]*

[1]Institute for Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, United States, [2]Department of Horticulture, University of Georgia, Athens, GA, United States

Different species of edible seed watermelons (*Citrullus* spp.) are cultivated in Asia and Africa for their colorful nutritious seeds. Consumer preference varies for watermelon seed coat color. Therefore, it is an important consideration for watermelon breeders. In 1940s, a genetic model of four genes, *R*, *T*, *W* and *D*, was proposed to elucidate the inheritance of seed coat color in watermelon. In this study, we developed three segregating F_2 populations: Sugar Baby (dotted black seed, *RRTTWW*) × plant introduction (PI) 482379 (green seed, *rrTTWW*), Charleston Gray (dotted black seed, *RRTTWW*) × PI 189225 (red seed, *rrttWW*), and Charleston Gray (dotted black seed, *RRTTWWdd*) × UGA147 (clump seed, *RRTTwwDD*) to re-examine the four-gene model and to map the four genes. In the dotted black × green population, the dotted black seed coat color (*R_*) is dominant to green seed coat color (*rr*). In the dotted black × red population, the dominant dotted black seed coat color and the recessive red seed coat color segregate for the *R* and *T* genes, where the *R* gene is dominantly epistatic to the *T* gene. However, the inheritance of the *T* locus did not fit the four-gene model, thus we named it T^1. In the dotted black × clump population, the clump seed coat color and the dotted black seed coat color segregate for *W* and *D*, where *D* is recessively epistatic to *W*. The *R*, T^1, *W*, and *D* loci were mapped on chromosomes 3, 5, 6, and 8, respectively, using QTL-seq and genotyping-by-sequencing (GBS). Kompetitive Allele Specific PCR (KASP™) assays and SNP markers linked to the four loci were developed to facilitate maker-assisted selection (MAS) for watermelon seed coat color.

Keywords: *Citrullus lanatus*, *Citrullus amarus*, edible seed watermelon, seed coat color, QTL-seq, KASP™ assay, SNP markers, epistasis

INTRODUCTION

Watermelon (*Citrullus lanatus*) is an annual, warm season vegetable crop which is grown throughout the tropical and sub-tropical regions of the world, predominantly for consumption of the sweet flesh. However, in many Asian and African countries, watermelons are instead cultivated for edible seeds. In China and India, most of the edible seed watermelons are from *C. lanatus* (Zhang, 1996; Mahla et al., 2014), whereas in West Africa, egusi watermelon, from the indigenous *C. mucosospermus* are extensively cultivated for edible seed (Oyolu, 1977; Gusmini et al., 2004). *C. colocynthis* is also cultivated as an edible seed watermelon in the Arabian peninsula and in India (Schafferman et al., 1998; Mahla et al., 2014). The land under edible seed watermelon

production is increasing and the market has expanded from China,
India and Africa to Europe and the Americas (Zhang, 1996; National Research Council, 2006; Mahla et al., 2014).

Seed coat color is an economically important trait because consumers prefer watermelon seeds with a specific color of seed. In China, seeds with red seed coat color, or seed with white center and a black margin are preferred (Zhang, 1996). Watermelon have a wide variety of seed coat colors ranging from flat black (solid black), dotted black (stipple black), tan, green, red, and clump to white (Poole et al., 1941; Poole, 1944). Flat black seeds have smooth, shiny, completely black seed coat, whereas dotted black seeds have a few to numerous black dots on an undercoat that can vary in color from whitish to red or even green. These black dots, which can usually be felt as protruding pins, provide dotted black seed coat a rough texture. Tan, green, and red seed coat have different shades of brown, green, and red color, respectively. Clump seed coat either have black pigment throughout the seed surface except on the narrow line inside the margin of the seed or have a white center, with a black rim on margin and/or two black spots on the hilum end (Poole et al., 1941). Description and naming of seed phenotypes has often been inconsistent among authors (Weetman, 1937; Poole et al., 1941; Sachan and Nath, 1976; Nath and Khandelwal, 1978) and for the sake of simplicity, we will use the phenotypic classification used by Poole et al. (1941).

Seed coat color is genetically controlled by a number of genes involving complex genetic interactions (Poole et al., 1941). The earliest attempt to study the inheritance of seed coat color was by Kanda (1931). He crossed flat black seeded watermelon with dotted black seeded watermelon and demonstrated that the flat black seed coat color is monogenically dominant to the dotted black seed coat color. Later, McKay (1936) developed two crosses: tan × red and green × red and showed that in each cross, the former phenotype was monogenically dominant to the latter phenotype. Weetman (1937) developed populations from three different crosses and showed that (1) the dotted black seed coat is monogenically dominant to the clump seed coat, and (2) different combinations of two genes produce clump and tan seed coat color. Poole et al. (1941) developed 40 different segregating populations and from the results they proposed a four-gene model controlling seed coat color in watermelon. According to this model, different combinations of three genes: R, T, and W with a modifier gene D (which only acts when the other three genes are in the dominant state) produce different seed coat colors, like flat black ($RTWD$), dotted black ($RTWd$), green (rTW), tan (RtW), clump (RTw), red (rtW), white tan-tip (Rtw), and white pink-tipped (rtw). This 1941 study was the last in-depth, large scale study on the genetics of watermelon seed coat color.

Next generation sequencing (NGS) technologies have made high-throughput sequencing less error-prone and very cost effective. As a result, NGS has become popular for the discovery of molecular markers throughout the genome (Varshney et al., 2009). GBS is a simple but highly scalable NGS-based genotyping model that can be used to genotype large populations and to identify thousands of genomic markers throughout the genome simultaneously (Elshire et al., 2011). GBS has been widely used to develop linkage maps and map quantitative trait loci (QTL) in several crops including watermelon (Lambel et al., 2014; Meru and McGregor, 2016a,b; Branham et al., 2017), zucchini (Montero-Pau et al., 2017), cucumber (Wang et al., 2018), pumpkin (Zhang et al., 2015), barley (Liu et al., 2014), pea (Boutet et al., 2016), rice (Bhatia et al., 2018), and alfalfa (Adhikari et al., 2018). Another relatively recent NGS-based technology is QTL-seq (Takagi et al., 2013). It combines bulk segregant analysis (Michelmore et al., 1991) with whole genome sequencing to identify QTL and to discover genetic markers necessary for MAS. One of the advantages of QTL-seq is that it does not require genotyping all the individuals in the population (Takagi et al., 2013). The first use of the QTL-seq approach in watermelon was to map a dwarfism locus on chromosome 7 (Dong et al., 2018). QTL-seq has been employed in several other crops like rice (Takagi et al., 2013), tomato (Illa-Berenguer et al., 2015), cucumber (Lu et al., 2014), chickpea (Das et al., 2015; Singh et al., 2016), and peanut (Clevenger et al., 2018).

Identification of the genomic regions associated with seed coat color is a crucial step in identifying candidate genes and in developing molecular markers for MAS. In this study, we used two interspecific and one intraspecific F_2 populations segregating for different seed coat colors to (1) determine the location of the R, T, W and D loci responsible for seed coat color development in watermelon and (2) determine the interaction among these loci.

MATERIALS AND METHODS

Plant Materials and Phenotyping

Three segregating F_2 populations were used to identify the loci responsible for seed coat color development in watermelon. The dotted black × green F_2 population ($n = 128$) was developed by crossing dotted black seeded Sugar Baby (*C. lanatus*) with green seeded PI 482379 (*C. amarus*) (**Figure 1A**). A dotted black × red F_2 population ($n = 96$) was developed by crossing dotted black seeded Charleston Gray (*C. lanatus*) with red seeded PI 189225 (*C. amarus*) (**Figure 1B**). The dotted black × clump population ($n = 178$) used in this study was developed by Meru and McGregor (2016b) to map *Fusarium oxysporum* f. sp. *niveum* race 2 in sweet watermelon. This F_2 population was produced by crossing dotted black seeded Charleston Gray (*C. lanatus*) with clump seeded UGA147 (*C. lanatus*), a selection from PI 169233 (**Figure 1C**).

The dotted black × green and dotted black × red parental cultivar/accession along with F_1 plants and both F_2 populations were sowed in the greenhouse on May 4, 2017 and transplanted in the field at the Durham horticulture farm (Watkinsville, GA) on May 30, 2017. The dotted black × green population was phenotyped in the field on August 24–25, 2017, and the dotted black × red population was phenotyped on September
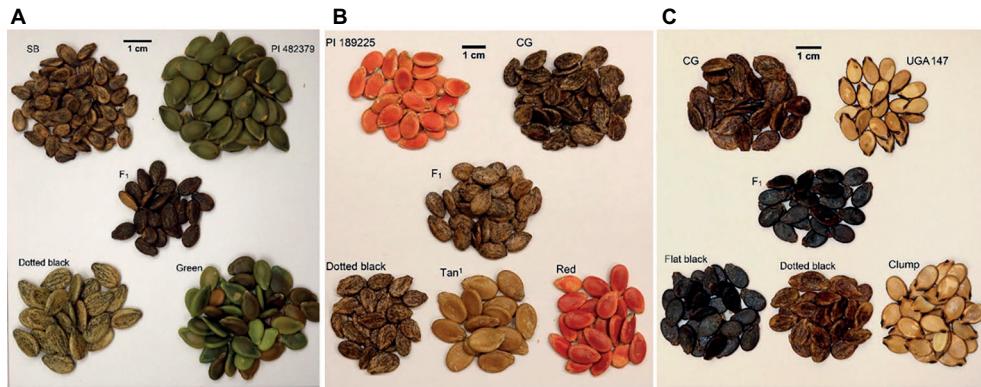
FIGURE 1 | Seed coat color of parents, F_1, and F_2 progenies in **(A)** the dotted black × green population, **(B)** the dotted black × red population and **(C)** the dotted black × clump population. In the dotted black × green population **(A)**, seed of dotted black seeded Sugar Baby (SB), female parent, green seeded PI 482379, male parent, F_1 and F_2 individuals with dotted black and green phenotype. In the dotted black × red population **(B)**, seed of red seeded PI 189225, female parent, dotted black seeded Charleston Gray (CG), male parent, F_1 and F_2 individuals with dotted black, tan^1 and red phenotype. In the dotted black × clump population **(C)**, seed of dotted black seeded Charleston Gray (CG), female parent, clump seeded UGA147, a selection from PI 169233, male parent, F_1 and F_2 progenies with flat black, dotted black and clump phenotype.

7–10, 2017. Mature fruits from each plant were cut open and seeds were visually phenotyped. Dry seeds from the parental, F_1 and F_2 plants from the dotted black × clump population, grown in the greenhouse in 2012 and 2013, were visually phenotyped under daylight conditions. Seeds were harvested between 40 and 48 days after pollination. For all populations, seed were classified as dotted black if black dots or stipples were observed that were rough to the touch, irrespective of the undercoat color. This is in line with the phenotype as described by Poole et al. (1941) when developing the four gene model.

Bulk Construction and DNA Isolation for QTL-seq

For QTL-seq of the dotted black × green population, a dotted black bulk (D-bulk) and a green bulk (G-bulk) were constructed by pooling equal amounts of DNA from 18 individuals of each phenotype. Similarly, for the QTL-seq of the dotted black × red population, a tan^1 bulk (T-bulk) was developed by pooling equal amounts of DNA from 20 individuals with tan^1 seed coat color and a red bulk (R-bulk) was developed by pooling equal amounts of DNA from 7 individuals with red seed coat color. Genomic DNA was extracted using the E. Z. N. A. Plant DNA kit (Omega Bio-Tek Inc., Norcross, GA) following the manufacturer's protocol. DNA concentrations were measured using an Infinite M200Pro plate reader (Tecan, Group Ltd., Mannerdorf, Switzerland), and bulks were comprised from equal amounts of DNA from the selected individuals and sent to the HudsonAlpha Institute for Biotechnology (Huntsville, AL) for library preparation and 151 bp paired-end whole genome sequencing on the Illumina HiSeq X (Illumina, San Diego, CA).

Analysis of NGS Data

A total of 168,613,320, 172,686,615, 124,764,246, and 154,206,455 reads for the D-bulk, G-bulk, T-bulk, and R-bulk

were generated from NGS, respectively. The quality of the reads obtained from NGS was analyzed using FastQC (Andrews, 2010). To ensure that the average phred score for all of the base positions in all the reads was higher than 28, bases with a low phred score were trimmed on both ends for all the bulks as follows: the first seven bases of forward and reverse reads of all bulks, the last two bases of all forward reads, the last 41 bases of reverse reads of the D-bulk, the last 27 bases of reverse reads of the G-bulk, the last 31 bases of reverse reads of the T-bulk, and the last 27 bases of reverse reads of the R-bulk. The downstream analysis for all the bulks was the same. The trimmed reads were aligned against the 97103 watermelon genome (Guo et al., 2013) using default BWA and BWA MEM options (Li and Durbin, 2009). 165,489,026 (98.15%) reads from the D-bulk, 170,855,441(98.94%) reads from the G-bulk, 123,169,517 (98.72%) reads from the T-bulk, and 152,346,070 (98.79%) reads from the R-bulk were aligned with an average depth > 83×. SAM tools (Li et al., 2009) were used to sort, index, and calculate the genotype likelihood. BCF tools and a custom-made python script were used for SNP calling and filtering. A total of 4,953,800 SNPs was identified between the D-bulk and the G-bulk, and 3,401,764 SNPs were identified between the T-bulk and the R-bulk. The SNP-index, which is the proportion of reads harboring SNPs divided by the total number of reads for a genomic position, was calculated for every base in the genome for all bulks. The SNP-index of the G-bulk was subtracted from the SNP-index of the D-bulk to obtain a ΔSNP-index for the dotted black × green population, and the SNP-index of the R-bulk was subtracted from the SNP-index of the T-bulk to obtain a ΔSNP-index for the dotted black × red population. A custom python script was used to conduct sliding window analysis by averaging the ΔSNP-index within a 1 Mb window region with a 10 kb stepwise incremental. A permutation test was conducted to develop a null model assuming no QTL as

explained by Takagi et al. (2013) and Clevenger et al. (2018). Thresholds for $p < 0.05$ and $p < 0.01$ were calculated for both population taking population size, number of individuals in each bulk, and read depth into account.

DNA Extraction of F₂ Populations and KASP™ Genotyping

Approximately 50 mg of leaf material from each individual of the dotted black × green and the dotted black × red parental cultivar/line, F₁, and F₂ populations were frozen in liquid nitrogen and ground using a TissueLyser II (QIAGEN, Hilden, Germany). DNA was extracted from leaf material using the King et al. (2014) extraction method with the following modifications. About 500 µl of extraction buffer mix [40% (v/v) 5 M NaCl and 60% (v/v) extraction buffer (200 mM Tris/HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS)] was added on ground leaf material. Samples were vortexed, incubated for 30 min at 60°C, and centrifuged for 10 min at 3600 rpm. An equal amount of supernatant and isopropanol was mixed and centrifuged to obtain DNA pellets. The DNA pellets were washed with 70% alcohol, dried, and resuspended in 200 µl TE buffer.

To validate the association of significant peaks with seed coat color, SNPs identified through QTL-seq were converted into KASP™ assays (**Table 1**). Primers were designed using Primer3Plus (Untergasser et al., 2007), and PCR amplification was done using a S1000™ Thermo Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). The 4 µl PCR reaction included 2 µl of 50–100 ng/µl genomic DNA, 1.96 µl 2× low rox KASP™ master mix (LGC Genomics LLC, Teddington, UK), and 0.06 µl primer mix for a final concentration of 0.81 µM. The PCR conditions for the KASP™ assays were set as follows: 95°C for 15 min, followed by 10 cycles of touch down PCR (95°C for 20 s, primer annealing temperature + 9°C for 25 s with 1°C decrease each cycle and 72°C for 15 s), then followed by 35 additional cycles (95°C for 10 s, primer annealing temp for 1 min, and 72°C for 15 s). PCR florescent end-point readings was done using an Infinite M200Pro plate reader (Tecan, Group Ltd.), and genotyping calls were carried out using KlusterCaller™ (LGC Genomics LLC). Individuals whose florescent end-point readings for markers were ambiguous were called missing data and were excluded from genotypic analysis. This caused discrepancies between the number of individuals that were phenotyped and genotyped.

Genotyping of Dotted Black × Clump and Construction of a Genetic Linkage Map

GBS of the dotted black × clump population is described in Meru and McGregor (2016b). The original 501 SNPs for the population was filtered using Joinmap 5.0 (Van Ooijen, 2006) for missing data (with up to 20% missing data) and segregation distortion ($p < 0.0001$). The remaining 230 SNPs were ordered using the regression mapping algorithm and grouped into linkage groups at LOD 5. A linkage map was generated using the Kosambi mapping function by converting recombination frequencies into map distances in centimorgan (cM).

RESULTS

Phenotypic Segregation in the Dotted Black × Green Population and Mapping of the *R* Locus

In the dotted black × green population, F₁ plants had seeds with dotted black seed coat color indicating that dotted black is dominant over the green (**Figure 1A**). Initially, it seemed that the F₂ progeny included dotted black, green, and brown seed phenotypes. However, upon closer inspection, it was established that the green seed turned brown over time and this difference was due to maturity. Green and brown seed could be observed in fruit harvested from a single plant (**Supplementary Figure S1**, **Figure 1A**). This phenotype was classified as green to conform with Poole et al. (1941). The F₂ progenies segregated at a ratio of 88 dotted black to 40 green seeded individuals. A chi-square goodness of fit test shows that the observed segregation ratio fits a 3:1 ratio ($\chi^2_{(0.05,1)} = 2.67$, $p = 0.10$). This result confirms the conclusion made by Poole (1944) that the dotted black ($R\_$) seed coat color is monogenically dominant to the green seed coat color (rr).

From QTL-seq, a significant ΔSNP-index peak ($p < 0.01$) was identified from 4.48 to 12.98 Mb on chromosome 3 of the *C. lanatus* genome (**Figures 2A,B**). A KASP™ assay, UGA3_5820134 (**Table 1**), was designed for a SNP located near the highest peak [5,820,134 bp on chromosome 3 of 97103 reference genome (Guo et al., 2013)] to test the association between the significant peak and the phenotype. The assay was able to accurately predict the phenotype of 85.7% ($n = 126$) of individuals (**Figure 2C**), confirming the association of this region with the *R* locus.

TABLE 1 | KASP™ assays used to test association between significant genomic regions, identified from QTL-seq, and the seed coat color phenotype in watermelon. The marker names indicate the chromosome number and physical position of the marker based on 97103 watermelon genome (Guo et al., 2013).

KASP assay	Primer type	Primer sequence (5′-3′)	Tm (°C)
UGA3_5820134	FAM	GAAGGTGACCAAGTTCATGCTTAGAGACACAAGAAAGTTGCAAAGG	61.1
	VIC	GAAGGTCGGAGTCAACGGATTTAGAGACACAAGAAAGTTGCAAAGT	61.1
	Reverse	TCATTTATTTCCCTCCTTAGCTTTCA	62.5
UGA5_4591722	FAM	GAAGGTGACCAAGTTCATGCTTTGTGAAATCAAAGATATGGACCAA	61.7
	VIC	GAAGGTCGGAGTCAACGGATTTTGTGAAATCAAAGATATGGACCAG	61.7
	Reverse	GAGTTACTTGAATTTGGAAAGGAAAGG	62.6
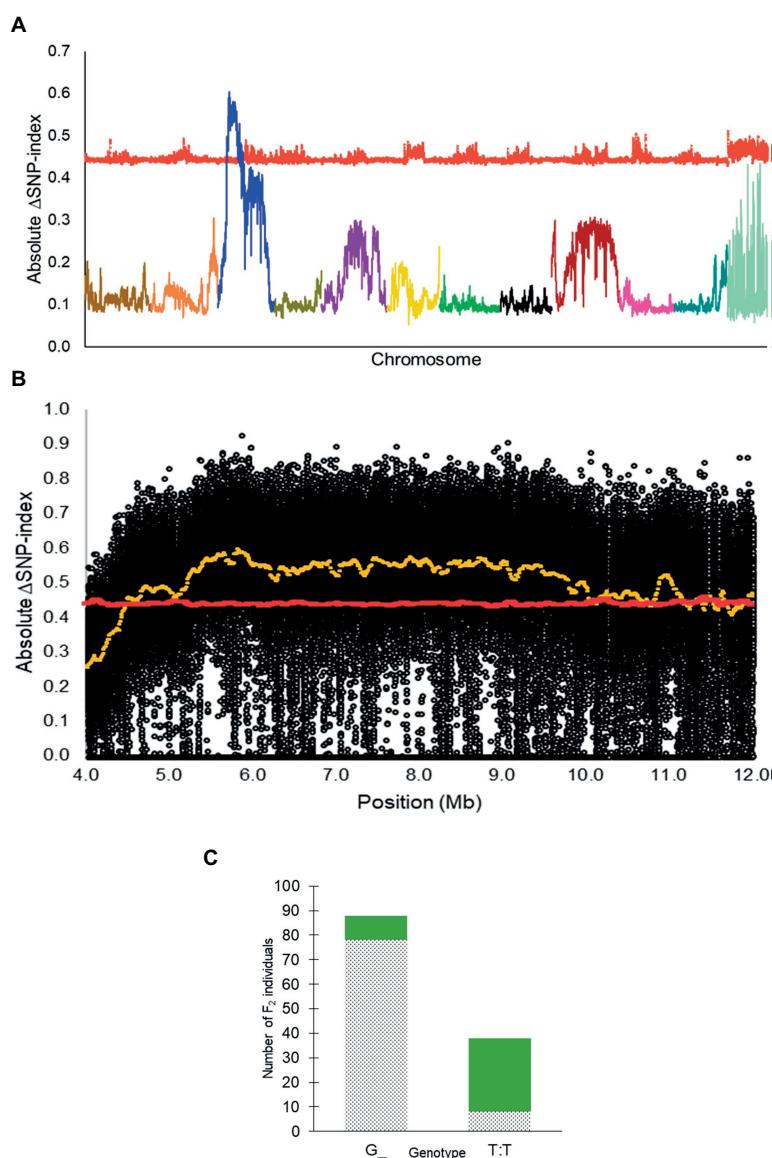
FIGURE 2 | (A) Absolute ΔSNP-index of all chromosomes in the inter-specific dotted black × green F$_2$ population developed from a cross between Sugar Baby (*C. lanatus*) and PI 482379 (*C. amarus*) plotted along with statistical confidence intervals under the null hypothesis of no QTL (*p* = 0.01) (red line). The *R* locus is mapped on chromosome 3. **(B)** Magnified view of *R* locus, a significant ΔSNP-index peak (yellow), along with absolute ΔSNP-index of SNPs (black circles) plotted against the SNP position. SNP positions are based on the 97103 watermelon genome (Guo et al., 2013). **(C)** Association of KASP™ marker UGA3_5820134 with seed coat color phenotype in the dotted black × green F$_2$ population (*n* = 126). The dotted black and green sections in the graph indicate the number of F$_2$ individuals with dotted black and green seed coat color, respectively.

Phenotypic Segregation in the Dotted Black × Red Population and Mapping of the *T¹* Locus

The F$_1$ plants in the dotted black × red population have seeds with dotted black seed coat denoting that the dotted black seed coat is dominant over the red seed coat color (**Figure 1B**). The segregating F$_2$ progenies had either dotted black, red or tannish (light shade of brown with yellowish tinge, similar to khaki) seed coat color. According to the four-gene model, F$_2$ individuals in a dotted black × red population is expected to have either dotted black, tan, green or red seed coat color at a 9 dotted black (*R_T_*): 3 tan (*R_tt*): 3 green (*rrT_*): 1 red (*rrtt*) ratio. In the current study, no individuals with green seed color were observed in the dotted black × red population. The tannish seed coat color observed was different from the range of brown color ("dark Tuscany brown to cacao") used to describe tan seed coat color by Poole et al. (1941). Therefore, we classified this phenotype as tan[1]. The F$_2$ progenies segregated at the ratio of 67 dotted black: 22 tan[1]: 7 red which statistically corresponds to 12:3:1 ratio ($\chi^2_{(0.05,2)}$ = 1.40, *p* = 0.49) and indicates dominant epistasis.
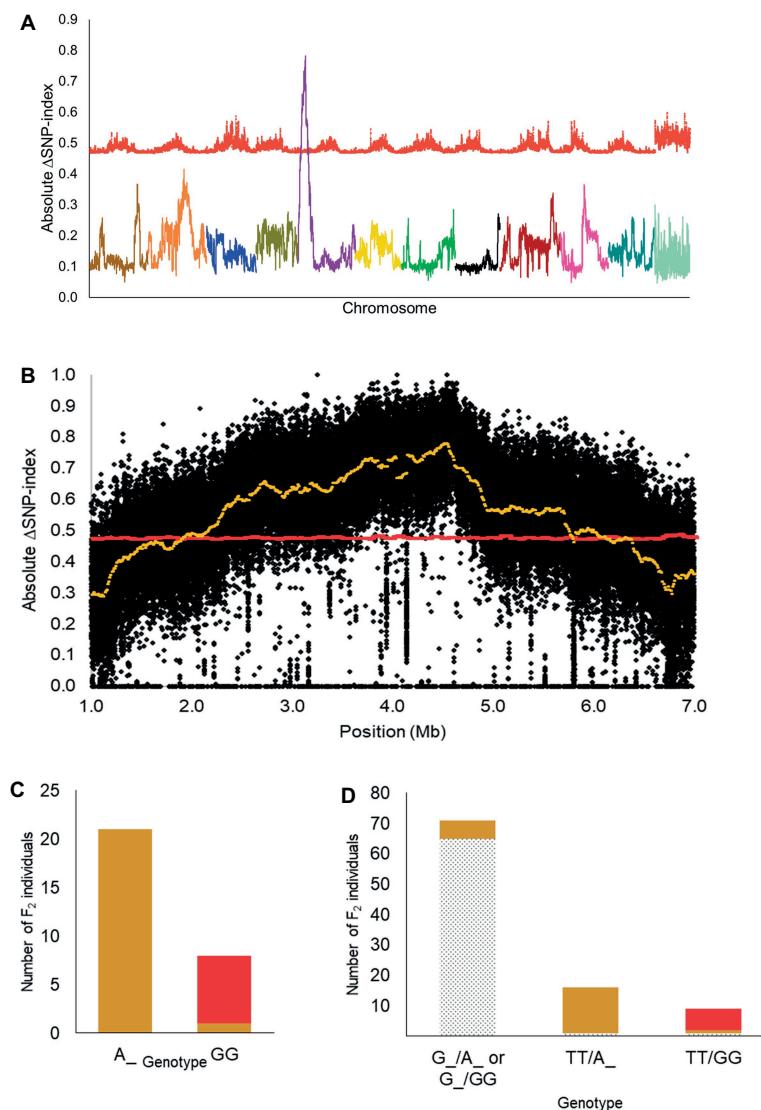
FIGURE 3 | (A) Mapping of the T^1 locus as a significant peak on chromosome 5 using QTL-seq. Absolute ΔSNP-index of all chromosomes, obtained by subtracting SNP-index of the red bulks from the tan[1] bulk in the dotted black × red population, is plotted along with statistical confidence intervals under the null hypothesis of no QTL (p = 0.01) (red line). Chromosomes are aligned in sequential order from 1 to 11 and 0. **(B)** Magnified view of significant ΔSNP-index peak (yellow) associated with T^1 locus along with absolute ΔSNP-index of SNPs (black circles) plotted against SNP position based on 97103 watermelon genome (Guo et al., 2013).
(C) Association of KASP™ marker UGA5_4591722 with the tan[1] and red seed coat phenotype in the dotted black × red F_2 population (n = 29). The x-axis denotes the genotype of F_2 individuals for KASP™ marker UGA5_4591722 and the y-axis denotes the number of F_2 individuals with tan[1] (tan bar) and red seed coat color (red bar). **(D)** Bar graph indicating the phenotypic prediction accuracy of KASP™ markers UGA3_5820134 and UGA5_4591722 in the dotted black × red population (n = 96). The genotypes on the x-axis represent the alleles of the UGA3_5820134/UGA5_4591722 markers. The dotted black, tan, and red sections in the graph indicate the number of F_2 individuals with respective seed coat color.

Based on the 12:3:1 ratio associated with dominant epistasis, we inferred that the tan[1] seed coat color and the red seed coat color were segregating for a single gene. Therefore, we pooled DNA from individuals with the tan[1] seed coat color and the red seed coat color to develop the T-bulk and the R-bulk, respectively. From QTL-seq, a significant ΔSNP-index peak (p < 0.01) was mapped from 1.89 to 6.46 Mb on chromosome 5 of the *C. lanatus* genome (**Figures 3A,B**). A SNP present within a significant peak and positioned at 4,591,722 bp on chromosome 5 of the 97103 reference genome

(Guo et al., 2013) was utilized to develop the UGA5_4591722 KASP™ assay (**Table 1**) to test the association of the peak and the phenotype. The marker was able to predict tan[1] (genotype: A:A or G:A) or red seed color (genotype: G:G) with 96.55% accuracy (n = 29) validating that the peak is related to the seed coat color (**Figure 3C**). Since the region mapped in this population was different from the region mapped in the dotted black × green population, we concluded that this region is not the *R* locus and based on the nature of inheritance, it can be inferred that this region is either a
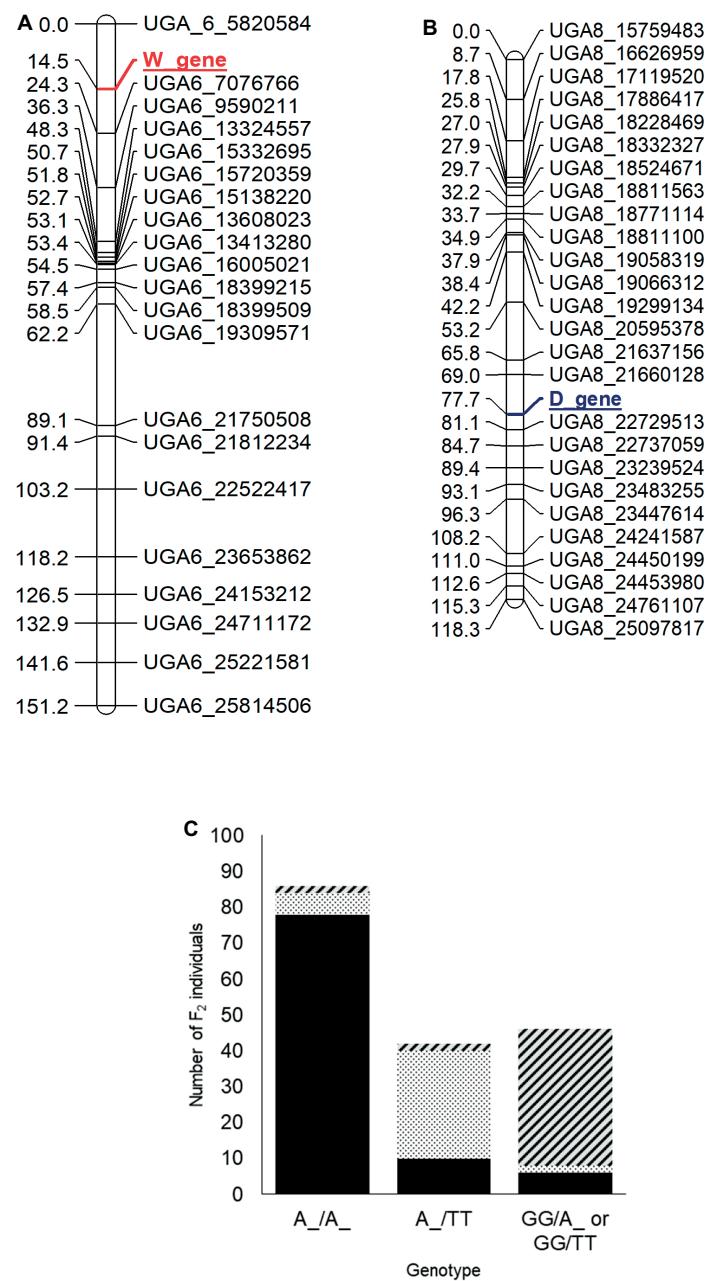
FIGURE 4 | Genetic maps of chromosome 6 **(A)** and 8 **(B)** showing the position of the W and D loci, respectively, in the dotted black × clump F_2 population developed from a cross between dotted black seeded Charleston Gray and clump seeded UGA147 (selection from PI 169233). The marker names indicate the chromosome number and physical position of the marker based on the 97103 watermelon genome (Guo et al., 2013). **(C)** Bar graph indicating phenotypic prediction accuracy of SNP markers UGA6_7076766 and UGA8_22729513 in the dotted black × clump population (n = 174). The genotype on the x-axis represents the alleles of markers UGA6_7076766/UGA8_22729513 in the F_2 population. The sections with black, dotted black, and diagonal black lines in the graph indicate number of F_2 individuals with flat black, dotted black and clump seed coat color, respectively.

novel locus or a different allele of the T locus described by Poole et al. (1941). Therefore, we propose to name this locus T^1, in conformance with gene nomenclature rules for Cucurbitaceae (Cucurbit Gene List Committee, 1982).

We also tested the KASP™ assay UGA3_5820134 associated with the R locus and found that the dotted black × red population was segregating for the R locus, as predicted by the four-gene

model. Approximately 97.01% of individuals with dotted black seed color had the genotype G:G or T:G and 79.31% of individuals with tan[1] or red seed color had the genotype T:T (**Supplementary Figure S2**). In addition, the genotypic data from the combination of KASP™ markers UGA3_5820134 and UGA5_4591722 were analyzed to understand the interaction between the two loci. Out of 71 F_2 individuals that had the genotype G:G or T:G

for marker UGA3_5820134, 65 individuals (91.54%) had dotted black seed coat color, independent of the UGA5_4591722 genotype (**Figure 3D**). Among 16 F_2 individuals that had the genotype T:T for marker UGA3_5820134 and A:A or G:A for marker UGA5_4591722, 15 individuals (93.75%) had tan[1] seed coat color. Similarly out of 9 F_2 individuals that had the genotype T:T for marker UGA3_5820134 and G:G for marker UGA5_4591722, 7 individuals (77.77%) had red seed coat color. This confirms our hypothesis that the *R* locus is dominantly epistatic to T^1 locus.

Phenotypic Segregation in the Dotted Black × Clump Population and Mapping of the *W* and *D* Loci

Based on the four-gene model, the dotted black genotype (*RTWd*) and the clump genotype (*RTwD* or *RTwd*) segregate either for the *W* gene or for both *W* and *D* genes. In the dotted black × clump population, the F_1 had flat black seed coat color (*W_D_*) meaning that the genotype of the clump parent, UGA147, is expected to be *RRTTwwDD* and the population is segregating for both the *W* and *D* genes (**Figure 1C**). The F_2 progenies segregated as flat black (*W_D_*, n = 94), dotted black (*W_dd*, n = 35), and clump (*wwD_* or *wwdd*, n = 49) which statistically fits a 9:3:4 ratio ($\chi^2_{(0.05,2)}$ = 0.91, p = 0.63), confirming the conclusion by Poole et al. (1941) that the *D* gene is recessively epistatic to *W*.

For mapping of the *W* and *D* genes, the seed phenotypes were translated into the "*abhcd*" genotype code format as described in the Joinmap® 4 manual (Van Ooijen, 2006). For the *W* locus, all individuals with non-clump seed coat color (flat black and dotted black, genotype: *W_*) were coded *d* (not clump parent genotype) and individuals with clump seed coat color (genotype: *ww*) were coded *b* (clump parent genotype). Similarly, for the *D* locus, all the individuals with flat black seed color (genotype: *D_*) were coded *c* (not dotted black parent genotype), and individuals with dotted black seed coat color (genotype: *dd*) were coded *a* (dotted black parent genotype). Individuals with the clump phenotype (genotype: *D_* or *dd*) were coded as missing data since the genotype of clump seeded individuals could not be predicted from the F_2 phenotype. The two phenotypic markers along with 230 SNP markers were used to construct a genetic map. Thirteen linkage groups with a total length of 1,226 cM and an average marker distance of 5.3 cM were developed for the 11 watermelon chromosomes (**Supplementary Figure S3**). The *W* locus was mapped at 14.5 cM on chromosome 6 between markers UGA6_5820584 and UGA6_7076766 (**Figure 4A**). The closest marker UGA6_7076766 is 9.8 cM away from the *W* locus. The genomic region associated with *W* locus partially overlapped with the major seed length QTL, *Qsl6*[M] (Prothro et al., 2012; Meru and McGregor, 2013). This is in accordance with the conclusion by Poole et al. (1941) that the *W* locus is linked with the *L* locus associated with seed length. The *D* locus was mapped between markers UGA8_21660128 and UGA8_22729513 at

position 77.7 cM on chromosome 8 on the dotted black × clump genetic map (**Figure 4B**). The closest marker, UGA8_22729513 is 3.4 cM away from the *D* locus.

We analyzed the genotypic data of SNP markers UGA6_7076766 and UGA8_22729513 to examine if the combination of *W* and *D* loci could predict seed coat color. Whenever F_2 individuals were homozygous dominant or heterozygous for both *W* (A:A or A:G genotype for marker UGA6_7076766) and *D* locus (A:A or A:T genotype for marker UGA8_22729513), 90.69% of individuals had flat black seed color (**Figure 4C**). Similarly, when F_2 individuals were homozygous dominant or heterozygous for *W* locus but recessive for the *D* locus (T:T genotype for marker UGA8_22729513), 71.42% individuals had dotted black seed coat color. However, when F_2 individuals were homozygous recessive for the *W* locus, (G:G genotype for marker UGA6_7076766), 82.60% of individuals had clump seed color irrespective of the *D* locus. In total, the phenotypic prediction accuracy of markers UGA6_7076766 and UGA8_22729513 when used as a proxy for *W* and *D* loci was 82.60% (n = 174). The percentage of inaccurate phenotype prediction (17.40%) is similar to the value of total recombination between SNP markers and respective loci (13.2 cM). Our result confirms that the genomic regions identified on chromosome 6 and 8 are associated with the *W* and *D* locus, respectively, and that the *D* locus is recessively epistatic to the *W* locus.

DISCUSSION

Genetic mapping of traits has always been a subject of interest to plant breeders. With the advent of NGS, the genotyping process has become fast, highly accurate, and relatively cheap. However, phenotyping still remains as a major bottleneck for efficient mapping of genetic traits. In watermelon, seed coat color can be phenotyped in major categories through visual analysis. Within these broad phenotypic categories described by Poole et al. (1941) and used in this study, other subtler variation was observed. This variation could be attributed to both genetic and non-genetic factors. One of the most important factors is the maturity of seeds. In our study, we found that maturity creates different variations in the phenotype. The number and size of black dots and the background color (color beneath black dots) of the dotted black seeds varied not only among individuals but also within the same individual, indicating that at least some of the variation within this phenotype can be attributed to non-genetic factors. Less mature seeds usually had very few fine dots on light brown background whereas mature seeds had many large dots on dark brown background. The individuals with green seed color also had different shades of green ranging from light green to dark green to brownish green depending upon the maturity of the seeds. Within the same individual, more mature fruit had brownish green seed color, while less mature

fruit had light green seeds (**Supplementary Figure S1**, **Figure 1A**). Maturity has also been identified as one of the non-genetic factors to influence phenotype by Poole et al. (1941). One of the possible ways to avoid effects of maturity would be to phenotype seeds of the same maturity stage. Avoiding the effect of maturity could possibly allow the use of quantitative measurements to phenotype seed coat color. However, in the current study, the *C. lanatus* × *C. amarus* populations were also segregating for maturity, leading to different F_2 progenies' fruits maturing at different rates. Fruits of early maturing individuals mature relatively sooner, and the flesh of those fruits starts to ferment and affect the phenotype (Poole et al., 1941). Whereas, in late maturing individuals, seeds are immature and not showing the mature phenotype. Additional research to develop a better phenotyping method which can avoid the effect of non-genetic factors in phenotyping is essential to understand the subtler phenotypes and identify additional loci involved in the genetics of watermelon seed coat color.

The four-gene model was developed by Poole et al. (1941) based on the inheritance of seed coat color in their populations and the populations developed by McKay (1936). Since the development of the four-gene model, only a few studies have been conducted to study inheritance of seed coat color. Nath and Dutta (1973) crossed a tan seeded individual with a red seeded individual and found that tan is monogenically dominant to red as predicted by the model. Similar studies by Sachan and Nath (1976) and Nath and Khandelwal (1978), crossing flat black seeded and tan (referred to as brown in the study) seeded individuals, also fit the four-gene model. However, the model was contradicted in a cross made by Nath and Khandelwal (1978) where flat black seed coat color was monogenically dominant to red seed coat color. In a similar study conducted by Sharma and Choudhury (1982), they found that flat black seed coat color and white seed color segregate only for one gene instead of two or three genes (depending on whether white seed is white-tan tip or white-pink tip) as predicted by the four-gene model. In the current study, the inheritance of the *R*, *W* and *D* loci fit the four-gene model, however, inheritance of the T^1 locus did not. The T^1 locus mapped in the dotted black × red population could be a different allele of the *T* locus or even a novel gene. Further testing of allelism is hampered by the lack of information about the identity of the red parental genotype used in the study by McKay (1936), which was used by Poole et al. (1941) in developing the four-gene model. The genotype is simply described as "citron," which is equivalent to *C. amarus*, but no further information is provided. The parental genotypes, "Peerless" and "Baby Delight," which produced red phenotype when crossed in Poole et al. (1941), are not currently available to replicate the cross. Nevertheless, findings from the current study and several others demonstrate that the four-gene model is incomplete and requires amendment.

Seed coat color is a complicated trait not only because of the number of genes involved in conferring phenotype but also because of the interactions among these loci. Understanding inheritance of the seed coat color phenotypes, the genetics and interaction of the different genes involved, identifying new genes, allelic variations and interactions among them requires developing, phenotyping and genotyping many populations. This is an arduous task to be done in one study because of the time, labor, and cost involved. The easier solution for this is to analyze and compare results of several studies and derive a consensus conclusion. However, the lack of the standard phenotypic descriptors makes it difficult to do so. In each study, the authors develop their own phenotyping methodology which makes it difficult to compare results among experiments (Weetman, 1937; Poole et al., 1941; Sachan and Nath, 1976; Nath and Khandelwal, 1978). This has been exacerbated by the fact that some of the lines/cultivars used in previous studies are no longer available to replicate the crosses. Since the phenotypic description developed by Poole et al. (1941) is the most detailed among any studies previously conducted, we propose that future studies related to seed coat color in watermelon should use the phenotypic description developed by Poole et al. (1941). Any new phenotypic class like tan[1] should only be used if it is distinct from the previously developed class or has a different inheritance pattern.

CONCLUSION

To conclude, this is the first study to map seed color gene loci in watermelon and to report SNP markers associated with these loci. Most of the prior research related to the genetics of watermelon seed color was carried out before the advent of molecular tools. In this study, we mapped the *R*, T^1, *W* and *D* loci on chromosomes 3, 5, 6 and 8, respectively, and developed markers UGA3_5820134, UGA5_4591722, UGA6_7076766, and UGA8_22729513 for MAS of seed coat color in watermelon. Further research is necessary to determine whether T^1 is a different allele or different locus than the previously described *T* locus. Moreover, identification of the T^1 locus indicates that there are additional genes/alleles that confer seed coat color in watermelon. Our results also open future research opportunities to fine map genomic regions and identify the genes conferring seed coat color and to identify functional markers for MAS of seed coat color in watermelon.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

LP conducted research and wrote the manuscript as a part of his MS research. JC guided LP on analysis of QTL-seq

data. CM conceived the project, guided research and data analysis, and revised the manuscript before submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00788/full#supplementary-material

REFERENCES

Adhikari, L., Lindstrom, O., Markham, J., and Missaoui, A. M. M. (2018). Dissecting key adaptation traits in the polyploid perennial Medicago sativa using GBS-SNP mapping. *Front. Plant Sci.* 9:934. doi: 10.3389/fpls.2018.00934

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed March 7, 2017).

Bhatia, D., Wing, R. A., Yu, Y., Chougule, K., Kudrna, D., Lee, S., et al. (2018). Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length. *Euphytica* 214:41. doi: 10.1007/s10681-018-2119-1

Boutet, G., Carvalho, S. A., Falque, M., Peterlongo, P., Lhuillier, E., Bouchez, O., et al. (2016). SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics* 17:121. doi: 10.1186/s12864-016-2447-2

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

Clevenger, J. P., Chu, Y., Chavarro, C., Botton, S., Culbreath, A., Isleib, T. G., et al. (2018). Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker assisted selection. *Front. Plant Sci.* 9:83. doi: 10.3389/fpls.2018.00083

Cucurbit Gene List Committee (1982). Update of cucurbit gene list and nomenclature rules. *Cucurbit Genet. Coop. Rep.* 5, 62–66.

Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Kumar, V., et al. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res.* 22, 193–203. doi: 10.1093/dnares/dsv004

Dong, W., Wu, D., Li, G., Wu, D., and Wang, Z. (2018). Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci. Rep.* 8, 1–7. doi: 10.1038/s41598-018-21293-1

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

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

Gusmini, G., Wehner, T., and Jarret, R. (2004). Inheritance of egusi seed type in watermelon. *J. Hered.* 95, 268–270. doi: 10.1093/jhered/esh031

Illa-Berenguer, E., Van Houten, J., Huang, Z., and Van Der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor. Appl. Genet.* 128, 1329–1342. doi: 10.1007/s00122-015-2509-x

Kanda, T. (1931). The inheritance of seed-coat colouring in the watermelon. *Jpn. J. Genet.* 7, 30–48. doi: 10.1266/jjg.7.30

King, Z., Serrano, J., Boerma, H. R., and Li, Z. (2014). Non-toxic and efficient DNA extractions for soybean leaf and seed chips for high-throughput and large-scale genotyping. *Biotechnol. Lett.* 36, 1875–1879. doi: 10.1007/s10529-014-1548-8

Lambel, S., Lanini, B., Vivoda, E., Fauve, J., Patrick Wechter, W., Harris-Shultz, K. R., et al. (2014). A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. *Theor. Appl. Genet.* 127, 2105–2115. doi: 10.1007/s00122-014-2363-2

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

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

Liu, H., Bayer, M., Druka, A., Russell, J. R., Hackett, C. A., Poland, J., et al. (2014). An evaluation of genotyping by sequencing (GBS) to map the *Breviaristatum-e* (*ari-e*) locus in cultivated barley. *BMC Genomics* 15:104. doi: 10.1186/1471-2164-15-104

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. *Theor. Appl. Genet.* 127, 1491–1499. doi: 10.1007/s00122-014-2313-z

Mahla, H. R., Singh, J. P., and Roy, M. M. (2014). *Seed purpose watermelon in arid zone.* (Jodhpur, Raj, India: Central Arid Zone Research Institute), 44.

McKay, J. (1936). Factor interaction in *Citrullus* seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. *J. Hered.* 27, 110–112.

Meru, G., and McGregor, C. (2013). Genetic mapping of seed traits correlated with seed oil percentage in watermelon. *HortScience* 48, 955–959. doi: 10.21273/HORTSCI.48.8.955

Meru, G., and McGregor, C. (2016a). Genotyping by sequencing for SNP discovery and genetic mapping of resistance to race 1 of *Fusarium oxysporum* in watermelon. *Sci. Hortic.* 209, 31–40. doi: 10.1016/j.scienta.2016.06.005

Meru, G., and McGregor, C. E. (2016b). A genetic locus associated with resistance to *Fusarium oxysporum* f. sp. *niveum* race 2 in *Citrullus lanatus*-type watermelon. *J. Am. Soc. Hortic. Sci.* 141, 617–622. doi: 10.21273/JASHS03890-16

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

Montero-Pau, J., Blanca, J., Esteras, C., Martínez-Pérez, E. M., Gómez, P., Monforte, A. J., et al. (2017). An SNP-based saturated genetic map and QTL analysis of fruit-related traits in zucchini using genotyping-by-sequencing. *BMC Genomics* 18:94. doi: 10.1186/s12864-016-3439-y

Nath, P., and Dutta, O. (1973). Inheritance of some seed characters in *Citrullus lanatus* Thunb. Mansf. *Indian J. Hortic.* 30, 388–390.

Nath, P., and Khandelwal, R. (1978). Genetics of seed-coat colour in *Citrullus lanatus* (Thunb.) Mansf. *Theor. Appl. Genet.* 53, 29–31. doi: 10.1007/BF00273133

National Research Council (2006). *Lost crop of Africa: Volume II: Vegetables.* (Washington, D.C: The National Academies Press).

Oyolu, C. (1977). A quantitative and qualitative study of seed types in egusi (*Colocynthis citrullus* L.). *Trop. Sci.* 19, 287–290.

Poole, C. (1944). Genetics of cultivated cucurbits. *J. Hered.* 35, 122–128. doi: 10.1093/oxfordjournals.jhered.a105364

Poole, C. F., Grimball, P. C., and Porter, D. R. (1941). Inheritance of seed characters in watermelon. *J. Agric. Res.* 63, 433–456.

Prothro, J., Sandlin, K., Abdel-Haleem, H., Bachlava, E., White, V., Knapp, S., et al. (2012). Main and epistatic quantitative trait loci associated with seed size in watermelon. *J. Am. Soc. Hortic. Sci.* 137, 452–457. doi: 10.21273/JASHS.137.6.452

Sachan, S. C. P., and Nath, P. (1976). "Inheritance of some qualitative characters in watermelon, *Citrullus lanatus* (Thunb.) Mansf" in *Proc. Third. Intn. Symp. Sub-tropic. Tropic. Hort.* (Bangalore, 1972), 99–105.

Schafferman, D., Beharav, A., Shabelsky, E., and Yaniv, Z. (1998). Evaluation of *Citrullus colocynthis*, a desert plant native in Israel, as a potential source of edible oil. *J. Arid Environ.* 40, 431–439. doi: 10.1006/jare.1998.0454

Sharma, R., and Choudhury, B. (1982). Studies on inheritance of fruit and seed characters in watermelon *Citrullus lanatus*. *Veg. Sci.* 9, 89–95.

Singh, V. K., Khan, A. W., Jaganathan, D., Thudi, M., Roorkiwal, M., Takagi, H., et al. (2016). QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. *Plant Biotechnol. J.* 14, 2110–2119. doi: 10.1111/pbi.12567

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. doi: 10.1093/nar/gkm306

Van Ooijen, J. (2006). *JoinMap 4: Software for the calculation of genetic linkage maps in experimental populations.* (Wageningen: Kyazma B.V.).

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

Wang, Y., Vandenlangenberg, K., Wen, C., Wehner, T. C., and Weng, Y. (2018). QTL mapping of downy and powdery mildew resistances in PI 197088 cucumber with genotyping-by-sequencing in RIL population. *Theor. Appl. Genet.* 131, 597–611. doi: 10.1007/s00122-017-3022-1

Weetman, L. M. (1937). Inheritance and correlation of shape, size and color in the watermelon, *Citrullus vulgaris* Schrad. *Res. Bull. - Iowa Agric. Home Econ. Exp. Stn.* 228, 224–256.

Zhang, G., Ren, Y., Sun, H., Guo, S., Zhang, F., Zhang, J., et al. (2015). A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). *BMC Genomics* 16:1101. doi: 10.1186/s12864-015-2312-8

Zhang, J. (1996). Breeding and production of watermelon for edible seed in China. *Cucurbit Genet. Coop. Rep.* 19, 66–67.

CmVPS41 Is a General Gatekeeper for Resistance to *Cucumber Mosaic Virus* Phloem Entry in Melon

Laura Pascual[1†], Jinqiang Yan[1‡], Marta Pujol[1,2], Antonio J. Monforte[3], Belén Picó[4] and Ana Montserrat Martín-Hernández[1,2*]

[1] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, C/Vall Moronta, Edifici CRAG, Bellaterra (Cerdanyola del Vallés), Barcelona, Spain, [2] Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Campus UAB, Bellaterra, Barcelona, Spain, [3] Instituto de Biología Molecular y Celular de Plantas (IBMCP), Universitat Politècnica de València (UPV)-Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain, [4] COMAV, Institute for the Conservation and Breeding of Agricultural Biodiversity, Universitat Politècnica de València (UPV), Camino de Vera s/n, Valencia, Spain

Melon production is often compromised by viral diseases, which cannot be treated with chemicals. Therefore, the use of genetic resistances is the main strategy for generating crops resistant to viruses. Resistance to *Cucumber mosaic virus* (CMV) in melon is scarcely described in few accessions. Until recently, the only known resistant accessions were Freeman's Cucumber and PI 161375, cultivar Songwhan Charmi (SC). Resistance to CMV in melon is recessive and generally oligogenic and quantitative. However, in SC, the resistance to CMV strains of subgroup II is monogenic, depending only on one gene, *cmv1*, which is able to stop CMV movement by restricting the virus to the bundle sheath cells and preventing a systemic infection. This restriction depends on the viral movement protein (MP). Chimeric viruses carrying the MP of subgroup II strains, like the strain LS (CMV-LS), are restricted in the bundle sheath cells, whereas those carrying MP from subgroup I, like the strain FNY (CMV-FNY), are able to overcome this restriction. *cmv1* encodes a vacuolar protein sorting 41 (CmVPS41), a protein involved in the transport of cargo proteins from the Golgi to the vacuole through late endosomes. We have analyzed the variability of the gene *CmVPS41* in a set of 52 melon accessions belonging to 15 melon groups, both from the spp *melo* and the spp *agrestis*. We have identified 16 different haplotypes, encoding 12 different CmVPS41 protein variants. Challenging members of all haplotypes with CMV-LS, we have identified nine new resistant accessions. The resistance correlates with the presence of two mutations, either L348R, previously found in the accession SC and present in other three melon genotypes, or G85E, present in Freeman's Cucumber and found also in four additional melon genotypes. Moreover, the new resistant accessions belong to three different melon horticultural groups, Conomon, Makuwa, and Dudaim. In the new resistant accessions, the virus was able to replicate and move cell to cell, but was not able to reach the phloem. Therefore, resistance to phloem entry seems to be a general strategy in melon controlled by *CmVPS41*. Finally, the newly reported resistant accessions broaden the possibilities for the use of genetic resistances in new melon breeding strategies.

Keywords: melon, genetic diversity, VPS41, *Cucumber mosaic virus*, Resistance, Phloem loading

INTRODUCTION

Melon (*Cucumis melo* L.) belongs to the family Cucurbitaceae and is one of the most productive crops in the world, with above 30 million tones/year (FAOSTAT http://www.fao.org/faostat/en/). *C. melo* has been traditionally divided into two subspecies defined by the pubescence of the ovary, spp. *melo* and spp. *agrestis* (Kirkbride, 1993), but most recent classifications use horticultural groups defined by vine, flowering, fruit traits, and geographic criteria. Pitrat (2017) described 19 horticultural groups including wild, feral, and domesticated melons: agrestis, kachri, chito, tibish, acidulus, momordica, conomon, makuwa, chinensis, flexuosus, chate, dudaim, chandalak, indicus, ameri, cassaba, ibericus, inodorus, and cantalupensis. These groups represent the broad phenotypical variability in agronomical traits, such as ripening, sugar accumulation, or fruit morphology displayed by the cultivars and landraces of this species. Most sources of resistance to viruses and pests identified so far belong to the acidulus and momordica groups from India and to the Far Eastern group of conomon, chinensis, and makuwa melons (those frequently referred to as conomon group) (Robinson and Decker-Walters, 1997; Blanca et al., 2012; Leida et al., 2015).

One of the most devastating plant viruses for melon is *Cucumber mosaic virus* (CMV), which produces typical mosaic in leaves and fruits and stunting plants. CMV is the type member of the *Cucumovirus* genus. It is a viral species with high sequence variability, resulting in large number of strains that can infect a broad range of plant species, including economically important crops, such as other main cucurbits (watermelon, cucumber, squash, and zucchini) as well as crops of the Solanaceae and Cruciferae families (Edwardson and Christie, 1991). On the basis of their sequence, CMV strains are divided into two subgroups [subgroup I (SG I) and subgroup II (SG II)] showing ~70% sequence homology between groups (Roossinck, 2001). Genetic resistances are the most successful way of preventing viral infections. However, modern commercial cultivars usually lack genetic resistances, and it is necessary to introgress them from landraces and wild accessions (Pitrat, 2008; Pitrat, 2017). Until recently, only a few melon genotypes, mostly from Asia, have been reported as resistant to CMV. The most frequently resistance sources used in different studies have been the Japanese Freeman's Cucumber (Karchi et al., 1975) and PI 161375, the Korean cultivar "Songwhan Charmi" (Con-SCKo) (from now on, SC) (Risser et al., 1977), classified as conomon and chinensis, respectively (Pitrat, 2017). Genetic studies show that, in both cases, resistance is oligogenic, recessive (Pitrat and Lecoq, 1980), quantitative (Dogimont et al., 2000), and also strain specific (Diaz et al., 2003). Other studies report resistance in several cultivars of the makuwa group (Pitrat and Lecoq, 1980; Hirai and Amemiya, 1989). Surveys carried out more recently have found other sources of resistance, mostly Indian cultivars of the momordica group but also some Iranian accessions (Dhillon et al., 2007; Fergany et al., 2011; Malik et al., 2014; Argyris et al., 2015). For most of them, the genetic control remains undetermined, and the strain specificity of these resistances was not reported. Therefore, the introduction of resistances to CMV in commercial cultivars is still challenging, and likely, the combination of genes/alleles from different sources would contribute to a broad-based resistance against these viruses.

The most studied resistance to CMV reported to date is that derived from the SC genotype. It is strain specific, recessive, and complex, controlled by at least three quantitative trait loci (QTLs) (Guiu-Aragonés et al., 2014). The major QTL is the gene *cmv1*, which by itself confers resistance against CMV strains of SG II (Essafi et al., 2009; Guiu-Aragonés et al., 2015). *cmv1* is also necessary for resistance to strains of SG I, but in this case, it is not sufficient and requires the contribution of the other two QTLs, as an example of the defense–counter defense established between pathogen and host (Guiu-Aragonés et al., 2014). The recessive resistance genes against viruses usually encode host proteins recruited by the virus to complete its cycle. Mutations in these genes may lead to resistance. Most recessive resistance genes identified encode either eukaryotic translation initiation factors (eIFs) or other factors involved in virus accumulation (for a review, see (Hashimoto et al., 2016). However, unlike previously reported recessive resistance genes, *cmv1* is involved in the transport of the virus and prevents systemic infection (Guiu-Aragonés et al., 2016). In the inoculated leaf of the resistant plant, the CMV strain LS (SG II) is able to move from cell to cell up to the vein but remains restricted to the bundle sheath (BS) cells and does not enter the phloem (Guiu-Aragonés et al., 2016). Therefore, the plant is resistant to systemic infection. However, the strain FNY (SG I) can overcome this barrier and enter the phloem. The viral protein that determines if the virus is transported to the phloem is the movement protein (MP). A CMV carrying the MP of LS cannot be transported to the phloem when inoculated into a plant carrying *cmv1*. However, a CMV clone carrying the MP from FNY will be translocated (Guiu-Aragonés et al., 2015). *cmv1*, therefore, is a gatekeeper that determines if the virus is either transported to the phloem to produce a systemic infection (susceptible plant) or remains restricted to the BS cells (resistant plant). This discrimination occurs depending on the MP and only in the BS cells, since *cmv1* does not affect the cell-to-cell movement in other cells of the inoculated leaves.

Map-based cloning of *cmv1* has shown that this gene encodes a vacuolar protein sorting 41 (CmVPS41) (Giner et al., 2017). This protein is normally involved in the transport of proteins in vesicles from the late Golgi to the vacuole as part of the "homotypic fusion and vacuole protein sorting" complex (Asensio et al., 2013; Pols et al., 2013). CMV might, therefore, recruit CmVPS41 through its MP to be transported to the phloem. The polymorphism L348R, present in SC, was reported as the causal mutation that restricts phloem entry in this accession. (Giner et al., 2017).

Here, we explore the possibilities of *CmVPS41* as target for resistance by sequencing it in 50 new melon genotypes. We have selected the accessions to represent 15 different horticultural groups covering an ample spectrum of melon diversity. We report 14 new haplotypes and 9 newly discovered resistant accessions, including Freeman's Cucumber. The resistance to CMV shown by these genotypes was demonstrated to be allelic with that of SC. Moreover, the resistance is manifested as a restriction in phloem entry, as it occurs in the previously reported SC genotype (Guiu-Aragonés et al., 2016). These new alleles represent new opportunities for diversifying the resistance to CMV derived from *CmVPS41* during breeding programs.

MATERIALS AND METHODS

Plant and Virus Material

Melon genotypes used were a Spanish "Piel de Sapo" inbred line (In-PsSp) (from now on PS), traditionally included in the inodorus group and now classified within the ibericus group (Pitrat, 2017), the Korean accession PI 161375, cultivar Songwhan Charmi (Con-SCKo) (from now on SC), belonging to the chinensis group, and 50 additional accessions listed in **Table 1**. These accessions represent many of the groups described by Pitrat 2017: cantalupensis, ameri, chandalak, dudaim, chate, flexuosus, conomon, makuwa, chinensis, momordica, acidulus, tibish, chito, and agrestis. We included at least one accession per group and included a higher number of accessions from the groups where resistances had been previously described (conomon, makuwa, chinensis). Most of the selected accessions were molecularly characterized previously (Leida et al., 2015) and classified in seven structure groups. The selection included accessions from all the structure groups detected, thus representing a wide range of melon diversity (**Table 1**; **Supplementary Table 1**).

CMV strain LS, belonging to SG II and provided by Dr. J. Diaz-Pendón (CSIC, EELM, Málaga, Spain), was used for inoculations.

Sequencing *CmVPS41*, Analysis, and Identification of Haplotypes and Proteins.

CmVPS41 genomic sequence (http://melonomics.cragenomica.es/) was used to design seven primers pairs in order to amplify the complete coding sequence (Giner et al., 2017), using Primer3 software (Untergasser et al., 2012) (http://bioinfo.ut.ee/primer3-0.4.0/) (**Supplementary Table 2**). PCR amplified

fragments were purified with sepharose columns and sequenced by capillary electrophoresis at Macrogen (Macrogen Europe, Amsterdam, The Netherlands). Sequences were analyzed with Sequencher® version 5.0 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI, USA (http://www.genecodes.com) and aligned with PS exons in order to detect polymorphisms. The obtained information was used to reconstruct *CmVPS41* complete coding sequence for each accession and to define haplotypes. Nucleotide sequences for each identified haplotype were translated with ExPASy translation tool (http://web.expasy.org/translate/). DNA or protein sequences were aligned using ClustalW (Thompson et al., 1994). Phylogenic trees from the nucleotide alignment were calculated employing neighbor-joining method, with Jukes–Center distance and 100 bootstraps. For proteins, we employed neighbor-joining method, with Tamura distance and 100 bootstraps. For the *CmVPS41* gene, haplotype networks were constructed using PopArt (Leigh and Bryant, 2015) employing the minimum spanning method (Bandelt et al., 1999) Effects of the polymorphisms in the CmVPS41 protein were predicted with Protein Variation Effect Analyzer (PROVEAN) [http://provean.jcvi.org/index.php, (Choi et al., 2012)] changes were considered deleterious when the predicted PROVEAN score was lower than -2.5. The Plaza database, version 3.0 dicots (http://bioinformatics.psb.ugent.be/plaza/), was used to identify orthologues of the *VPS41* gene in other plant species whose genome was available.

Inoculation With CMV And Virus Detection

Inoculation and virus detection were done as previously described by Essafi and collaborators (2009), using at least six plants by tested genotype. Briefly, seeds were pregerminated by soaking them in

TABLE 1 | Accessions analyzed in this work. Indicated are CmVPS41 gene haplotype, CmVPS41 protein allele, response to the inoculation with CMV-LS number of accessions tested, accessions code, assigned group according to previous structure analysis.

Haplotype gene	Protein type (12)	CMV-LS*	Accessions**	Structure group***
Hap-1	Prot-1	S (2/4)	Ib-PSSP, Am-Bol, Can-VedFran, Con-YaPuJa	3, A, 1, (1,6)
Hap-2	Prot-2	S (8/26)	26 accessions****	5, (7,6), (6,2), A, (6,5), 6, (6,3), (6,7,2), (5,4)
Hap-2-C1012Irak	Prot-2-C1012Irak	S	Dud-C1012Irak	–
Hap-2-Ksud	Prot-2	–	Tibish-Ksud	7
Hap-2-QPMAfg	Prot-2	R	Dud-QPMAfg	6,7,2
Hap-3	Prot-3	R (4/4)	Con-SCKo, Con-GMJa, Con-Chi52Chi, Con-MielChi	6, (1,6), (6,5)
Hap-4	Prot-4	R (4/4)	Con-FreeCJa, Con-NanChi, Con-OgonChi, Con-ShiroJa	6
Hap-4-Pat81Ko	Prot-4-Pat81Ko	R	Con-Pat81Ko	6
Hap-5	Prot-5	S (2/2)	Flex-AryaInd, Ag-TriInd	A
Hap-5-Kakru	Prot-5	S	Ag-KakInd	–
Hap-5-TGR1551Zimb	Prot-5 TGR1551Zimb	S	Ac-TGR1551Zimb	7,6
Hap-6	Prot-2	–	Ag-15591Gha, Ag-TayCam	7
Hap-6-C38Nig	Prot-6-C38Nig	S	Ag-C38Nig	7
Hap-6-C836CV	Prot-6-C836CV	S	Ag-C836CV	–
Hap-7-CarBltA	Prot-7-CarBlta	S	Chate-CarBlta	1,4,5
Hap-7-SanIIPhil	Prot-7-SanIIPhil	S	Con-SanIIPhil	6,7

*Response to LS inoculation. S, susceptible; R, resistant. In parentheses, number of accessions tested out of total number of accessions with a given haplotype.
**First part of accession code indicates melon type, last part country of origin.
***Structure groups according to Leida et al. (2015): 1 and 2: Cantalupensis accessions, French charentais and reticulatus, respectively, 3: Spanish inodorus, 4: Worldwide inodorus and ameri, 5: Asian ameri, 6: Oriental melons, conomon, chinensis, and makuwa, 7: African agrestis, A: Admixture, accessions that could not be associated to any of the groups.
****Con-LongtChi, Con-KNMJa, Mom-KhaInd, Dud-DudaimFra, Con-CUM206Chi, Mom-PI124Ind, Chi-VellInd, Dud-QAPMSwitz, Mom-PI414Ind, Con-GouChi, Con-Baish1Chin, Con-Baish2Chin, Con-LongtChi, Con_Xiao1Chin, Con-HerChin, Am-KizilUzbe, Con-BaishChin, Con-Co6Chi, Con-GapPhi, Con-PauPol, Con-XiaoChi, Dud-DudaimAfg, Mom-MR1Ind, Ag-WCHInd, Am-ChanRus, Con-OmGMJa.

water overnight and then kept for 2–4 days in 12 h light at 28°C. Seedlings were grown in growth chambers (Sanyo MLR-350H, Sanyo Electric Biomedical Co, Osaka, Japan) during the whole essay, in long-day conditions of 22°C for 16 h with 5,000 lx of light and 18°C for 8 h in the dark. Viral inocula were prepared from freshly symptomatic leaves of zucchini ("Chapin F1," Semillas Fito SA, Barcelona) and rub inoculated onto the cotyledons of young melon plants still without leaves. Symptoms were scored visually at 7, 14, and 21 days postinoculation (dpi), and viral detection was done by double antibody sandwich ELISA in young developed leaves from the six inoculated plants. Analysis was done with CMV-specific polyclonal antisera (Loewe Biochemica GmbH, Otterfing, Germany) following manufacturer's protocol. ELISA reactions were measured spectrophotometrically at 405 nm using the VICTOR3 V multilabel plate reader (PerkinElmer). ELISA was considered positive when absorbance at 405 nm was larger than twice the negative control value. PS and SC were used as susceptible and resistant controls, respectively. SC12-1-99 NIL, derived from the NIL SC12-1 (Essafi et al., 2009), carrying a shorter introgression of SC that contains the *cmv1* gene, was also used as resistant control. Negative controls were prepared from leaf tissue extracts of mock-inoculated plants.

Crosses And Allelism Tests

A subset of the resistant accessions identified (at least one by haplotype) was crossed with melon accessions PS and NIL SC12-1-99. At least six F1 plants from each cross were tested against CMV-LS. After inoculation with CMV-LS, a resistance source was considered allelic only when none of the F1 plants from the cross with SC-12-1-99 plants showed symptoms, and the F1 plants from the cross with PS were infected.

Detection Of CMV-LS In The Phloem

First true leaf of six plants per genotype was inoculated with CMV-LS. After detection of yellow areas around the entrance points, the petiole of this leaf was collected at 11 dpi. RNA was isolated, and the presence of the virus in the phloem was tested by reverse transcription PCR as described (Guiu-Aragonés et al., 2015). CMV-LS, specific primer LS1-1400R (GAAGCATTCCACATATCGTAC), was used for RT, and the same primer together with LS1-900F (GTTTTATTTACAAG AGCGTACG) were used to amplify a 500-bp fragment of the viral genome.

RESULTS

Identification Of New CmVPS41 Melon Haplotypes And Protein Variants.

To study the genetic variability of the melon VPS41gene, 52 melon genotypes, including PS and SC, from 15 different melon groups were chosen. The *CmVPS41* genomic sequence was either sequenced or was obtained from the available whole genome sequences (Sanseverino et al., 2015) [those from the two controls, PS and SC, and from one cantalupensis melon, Can-VedFran, one

dudaim melon, Dud-C1012Irak, and one wild African *agrestis*, Ag-C836CV (**Table 1**)].

In the whole set of 52 accessions, 27 single nucleotide polymorphisms (SNPs) were identified (**Figure 1**), being 18 of them singletons and defining 14 new *CmVPS41* haplotypes (**Figure 1**), different from those from PS (Hap-1) and SC (Hap-3) (**Figure 2**). The most represented haplotype (Hap-2) was present in 26 melon genotypes (**Table 1**, **Figure 2B**). After removing the singletons, seven core haplotypes remained, and two haplotype blocks were evident, one from position 1 to 1,858 and another from 1,858 to 2,583 (**Table 1**, **Figure 1B**).

The haplotype 1 (Hap-1) was present in the two modern European cultivars, PS (the susceptible control) and Védrantais (Can-VedFran), representing the Spanish inodorus and French cantalupensis groups, respectively (**Table 1**). Hap-1 was also shared by two cultivars: the Japanese Yamato Purinsu (Con-YaPuJa), classified as makuwa, but grouped within the cantalupensis group by Leida et al. (2015), and the American landrace Am-Bol, from Bolivia, which also showed genome admixture in previous analysis (**Table 1**).

The haplotype 2 (Hap-2) was present in the vast majority of accessions, mostly from Asia, belonging to different horticultural types, such as the assayed Indian momordica, the chandalack and ameri accessions from Central Asia, and most of the makuwa and chinensis Far eastern melons (all in the well-defined structure group of oriental melons) (**Table 1**). Asian dudaim melons (molecularly related to the oriental group, but with admixture from other groups) also showed this haplotype, although two of the dudaim accessions from the Near East had unique Hap-2 derived haplotypes (Dud-C1012Irak and Dud-QPMAfg). Hap-2 was also present in two small-fruited accessions from India, the chito-type Velleri (Chi-VellInd) and the wild agrestis (Ag-WChInd). The only African type with Hap-2 was the Tibish accession Tibish-KSud, considered a melon domesticated in Africa from wild African agrestis (Endl et al., 2018) (**Table 1**, **Figure 2B**).

The other five haplotypes likely arose by single nucleotide mutations from haplotype 2 (**Figure 2B**) mostly in positions between 254 and 1,043, while no mutation was observed between positions 1,858 to 2,583, suggesting that the SNPs in this part of the sequence are in high linkage disequilibrium, whereas SNPs in the first part of the gene are in linkage equilibrium.

Interestingly, two of these five haplotypes, haplotypes 3 (Hap-3, that of the resistant control SC) and 4 (Hap-4), were present in the remaining accessions of the oriental melon group from China, Korea, and Japan (conomon, chinensis, and makuwa) that belong to the same structure group than Far Eastern melons displaying Hap-2 (**Table 1**).

Haplotype 5 (Hap-5) was found in two wild Indian agrestis, trigonus and kakru (Ag-TriInd and Ag-KakInd), and in the unique accessions analyzed of the flexuosus (Flex-AryaInd) and acidulus (Ac-TGR1551Zimb) groups, whereas haplotype 6 (Hap-6) was characteristic of the wild Central African agrestis accessions (**Table 1**). The last haplotype (Hap-7) was found in two accessions, one Italian (Chate-CarBIta) and one chinensis type from the Philippines (Con-SanIlPhi) (**Table 1**).

A

nt position	Hap - 1	Hap - 2	Hap-2-C1012Irak	Hap - 2 -Ksud	Hap - 2 -QPMAfg	Hap - 3	Hap - 4	Hap - 4 -Pat81Ko	Hap - 5	Hap - 5 -Kakru	Hap - 5-TGR1551Zimb	Hap - 6	Hap - 6-C38Nig	Hap - 6-C836CV	Hap - 7-CarBItA	Hap - 7-SanIlPhil
65-66	-	-	-	-	-	-	-	-	-	-	GGA	-	-	-	-	-
246	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C
254	G	G	G	G	G	G	A	A	G	G	G	G	G	G	G	G
451	A	A	A	A	A	A	A	A	T	T	T	A	C	A	A	A
478	C	C	C	C	C	C	C	C	C	C	C	T	T	T	C	C
720	T	T	T	T	T	T	T	T	T	T	T	T	G	T	T	T
784	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
925	C	C	C/T	C	C	C	C	C	C	C	C	C	C	C	C	C
1043	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T
1080	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C
1140	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G
1263	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A
1527	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	C
1858	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1924	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T
2139	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A
2259	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C
2451	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2475	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C
2560	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T
2583	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2758	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C
2781	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C
2822	C	C	C	C	C	C	C	C	C	C	C	C	C	C	A	A
2852	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T
2865	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T

B

nt position	PS	Am-ChanRus	SC	Con-FreeCJa	Flex-AryaInd	Ag-15591Gha	Chate-CarBIta
	Hap - 1	Hap - 2	Hap - 3	Hap - 4	Hap - 5	Hap - 6	Hap - 7
254	G	G	G	A	G	G	G
451	A	A	A	A	T	A	A
478	C	C	C	C	C	T	C
784	C	G	G	G	G	G	G
1043	T	T	G	T	T	T	T
1858	T	C	C	C	C	C	C
2451	G	A	A	A	A	A	A
2583	A	G	G	G	G	G	G
2822	C	C	C	C	C	C	A

FIGURE 1 | Haplotypes and frequency observed in the *CmVPS41* gene from 52 melon genotypes. **(A)** Nucleotide polymorphisms present in the melon haplotypes analyzed. In gray, the polymorphisms found in the sequenced accessions. **(B)** Core haplotypes excluding singletons. nt: nucleotide number.

The New CmVPS41 Haplotypes Produce 10 New Variants.

Out of the 27 polymorphisms found (including the two changes in position 451), 15 produced synonymous substitutions and one added an amino acid. The 11 nonsynonymous substitutions found produced amino acid changes with different effects on the final protein (**Table 2**, **Figure 3**). Eight of those substitutions have effects catalogued as neutral for the function of the protein. However, two of them had a strong theoretical deleterious effect, the G85E substitution, with a PROVEAN score of -7.008, and the L348R substitution, with a PROVEAN score of -5.929, whereas the D509K had a weak, almost neutral effect of -2.857.

The L348R substitution is the previously reported in SC and identified as the causal mutation of the resistance to

FIGURE 2 | (A) *CmVPS41* phylogenetic tree of the different haplotypes. Branches where all the accessions showing a given haplotype are resistant are marked in red. **(B)** Reconstruction of haplotypes evolution in a network. Polymorphisms are indicated as vertical lines, black synonymous SNPs. Green, amino acid change with neutral effect according to the PROVEAN score. Red, amino acid change with deleterious effect. Indels are not analyzed. Each circle represents a different haplotype, and circle size is proportional to the number of accessions carrying the haplotype. Circles are colored by melon type. Haplotypes were all the accessions are resistant are written in red.

TABLE 2 | Description of the amino acid changes predicted in the CmVPS41 protein. The effect according to Protein Variation Effect Analyzer (PROVEAN) is indicated.

nt position	Hap1 codon	Alternative	aa change	Shift value	Effect
65-66	GAC	GAGGAC	D22ED	3.231	Neutral
254	GGG	GAG	G85E	-7,008	Deleterious
451	ACT	TCT/CCT	T151S/T151P	0,889/-1.852	Neutral/neutral
784	CCA	GCA	P262A	-0,903	Neutral
925	CCT	TCT	P309S	-2,254	Neutral
1,043	CUA	CGA	L348R	-5,929	Deleterious
1,527	AAC	AAG	N509K	-2.857	Deleterious
1,858	TCA	CCA	S620P	1,203	Neutral
2,560	TCT	ACT	S854T	-0,464	Neutral
2,822	ACA	AAA	T941K	-0,545	Neutral
2,852	ATT	AGT	I951S	-0,328	Neutral

CMV-LS exhibited by this genotype (Giner et al., 2017). It is also present in other three genotypes (those sharing the Hap-3), all belonging to the makuwa and chinensis groups: the Japanese cultivar Ginsen makuwa (Con-GMJa), and the Chinese cultivars Miel Blanc (Con-MielChi) and China51 (Con-Chi51Chi) (**Table 1**). The latter was classified within the French cantalupensis group by Leida et al. (2015), but showed a high degree of admixture with the oriental melons group to which the other cultivars belong. The substitution G85E has not been previously reported and is present in five Far Eastern accessions, belonging to the conomon, makuwa, and chinensis groups (oriental melons structure group), which

share Hap-4. Two of the cultivars showing this mutation are the conomon Japanese cultivars Freeman's Cucumber (Con-FreeCJa) and Shiro uri Okayama (Con-ShiroJa), as well as two Chinese makuwa [Nanbukin (Con-NanChi) and Ogon9 (Con-OgonChi)] and one Korean chinensis Pat 81 (Con-Pat81Ko). The accessions sharing the same amino-acid sequence either in Hap-3 or in Hap-4 represent true different accessions, as they show clear phenotypic and molecular polymorphism (Leida et al., 2015; **Supplementary Figure 1**). Therefore, the haplotype was fixed before the diversification of these cultivars.

Our results indicate that the most common *CmVPS41* haplotype (Hap-2) is not restricted to specific melon groups,

FIGURE 3 | Polymorphism in *CmVPS41*. **(A)** *CmVPS41* genomic sequence. Boxes are exons. Polymorphisms with respect to In-PsSp *CmVPS41* are marked. Black, synonymous substitutions. Green, amino acid change with neutral effect. Red, amino acid change with deleterious effect. **(B)** Secondary structure of CmVPS41 protein from PS. In blue, β sheets; in red, α helix. All amino acid changes are indicated. Green, change with neutral effect. Red, change with deleterious effect. **(C)** Phylogenetic tree of the different protein alleles. In red, resistant genotypes. In blue, susceptible genotypes. The relative effect of their amino acid changes is shown by a gradient of red color, from neutral to deleterious, according to the values given in **Table 2**.

being present in highly diverse Asian accessions from different origins and representing different horticultural groups. However, some less frequent haplotypes only occur or are more frequent in specific groups, such as Hap-3 and Hap-4 in oriental melons and Hap-5 and Hap-6 in wild Indian and African agrestis, respectively.

Only Melon Genotypes Carrying CmVPS41 Variants With Strong PROVEAN Score Are Resistant To CMV-LS

Twenty-nine melon genotypes belonging to all melon groups and showing different *CmVPS41* haplotypes were challenged with CMV-LS to assess their resistance to this strain. As shown in **Table 1**, out of the 29 genotypes inoculated, 10 were resistant to CMV-LS. Nine of them carry the amino acid substitutions L348R (Ginseng makuwa, China 51, Miel Blanc, as well as the previously described SC) or G85E (Freeman's Cucumber, Shirouri Okayama, Pat81, Nanbukin, and Ogon9), whereas none of the genotypes carrying other amino acid substitutions correlated with resistance to CMV-LS. L348R and G85E are the polymorphisms showing the strongest PROVEAN score (**Figure 3**).

The role of L348R substitution on the resistance found in SC has previously been demonstrated (Giner et al., 2017). Therefore, this substitution should also be responsible for the resistance of the three new carrier genotypes. Regarding the substitution G85E, reported here for the first time, it is shared by Pat81 and the other four genotypes with Hap-4, Freeman's Cucumber, Shiro

Uri Okayama, Nanbukin, and Ogon 9. Pat81 carries an additional acid change with respect to PS, I951S, that shows a very weak PROVEAN score on the protein. As the mutation G85E is shared with Freeman's Cucumber and the other three resistant genotypes, which lack any other mutation, it is reasonable to think that the polymorphism responsible for resistance in Pat81 is G85E, whereas I951S has no effect on the phenotype. The substitution N509K reveals a deleterious, but almost neutral score, and indeed, the only accession carrying it, an Italian Landrace belonging to the ancient group of chate melons (Sabato et al., 2019) (Chate-CarBIta, haplotype 7-CarBItA), is susceptible to CMV-LS (**Table 1**, **Figure 3**). Therefore, these observations suggest that only those amino acid variations with a strong PROVEAN score lead to resistance to CMV-LS.

Interestingly, in the assayed collection, the dudaim cultivar Queen's pocket melon (Dud-QPMAfg) (Hap- 2) do not present any amino acid change with a strong effect on the protein and is, nevertheless, resistant to CMV-LS. QPM presents the most frequent amino acid sequence. Out of the 29 genotypes analyzed, showing this amino acid sequence, 9 of them were tested for resistance to CMV-LS, and only this dudaim cultivar was resistant to CMV-LS (**Table 1**), which may indicate that either the resistance from this accession has another genetic control or the phenotype is due to any other change, linked to 1,263 nucleotide synonymous mutation, not included in the current report.

Therefore, we have characterized the variability of the gene *CmVPS41* in 50 new genotypes, 9 of them resistant to CMV-LS

which belong to 4 melon botanical groups: Ginsen makuwa, China 51, Nanbukin and Ogon (makuwa), Miel blanc and Pat 81 (chinensis), Freeman's Cucumber and Shiro Uri Okayama (conomon), and Queen's pocket melon (dudaim), which could be used in breeding programs, depending on the phenotype or the need for additional resistances.

Allelism Tests

To confirm that *CmVPS41* was responsible for the resistant phenotype in the newly reported melon accessions, plants from Ginsen makuwa, Pat81, Freeman's Cucumber, and Queen's pocket melon were crossed with the near isogenic line SC12-1-99, which contains a small introgression from SC carrying the *CmVPS41* gene (Hap-3), in the PS background (Essafi et al., 2009). The same accessions were crossed with the susceptible genotype PS as control. All PS F1 hybrid plants were infected, confirming that the resistance in these accessions is recessive, as it is in SC (Pitrat and Lecoq, 1980; Guiu-Aragonés et al., 2014). However, each F1 progeny from crosses with SC12-1-99 was resistant to CMV-LS systemic infection. Given that the resistance is recessive, this result confirms that in all tested accessions, it is controlled by the same gene. Interestingly, the F1 from Queen's pocket melon x SC12-1-99 is also resistant to CMV-LS. Therefore, it is allelic also in this genotype, which implies that *CmVPS41* from Queen's pocket melon is responsible for the resistance to CMV-LS despite its predicted protein (Prot-2) being equal to other susceptible accessions (**Table 1**).

The New Resistant Genotypes Restrict CMV-LS Phloem Entry

Resistance to CMV-LS in the accession SC acts at the level of viral phloem entry (Guiu-Aragonés et al., 2016), whereas the virus can replicate and move cell to cell. To test if in the new resistant accessions CMV-LS was able to invade the phloem, the first true leaf of plants from accessions Ginsen makuwa, Pat81, Freeman's Cucumber, and Queen's pocket melon was inoculated with CMV-LS. After 11 dpi, all accessions had developed yellow areas around the entrance points in the inoculated leaf, indicating that the virus was able to replicate and move cell to cell in these accessions (**Figure 4A**), as had been previously reported for SC (Guiu-Aragonés et al., 2016). However, none of the accessions had developed symptoms of CMV infection except PS, used as positive control. Then, the petiole of the inoculated leaf was collected, and the presence of the virus in the phloem was tested by reverse transcription PCR. As shown in **Figure 4B**, the virus was absent from the phloem of all resistant accessions, including the resistant control SC, whereas in PS, the virus had reached the phloem and could be detected in the petiole of the inoculated leaf. This indicated that the resistance to CMV-LS in the four tested accessions was acting at the level of phloem entry, as is also the case of SC, where the virus is restricted in the BS cells of the minor veins (Guiu-Aragonés et al., 2016). Therefore, restriction to phloem entry seems to be a common strategy for recessive resistance to CMV in melon.

DISCUSSION

We have explored the polymorphism of the gene *CmVPS41* in 52 melon genotypes belonging to a representation of most major horticultural groups reported in the species (Pitrat, 2017). In total, 16 haplotypes have been observed, including PS and SC, leading to 12 different protein variants, including those already described from PS and SC (Giner et al., 2017). VPS41 is a protein involved in intracellular trafficking of proteins and vesicles from late Golgi to the vacuole that seems to have an essential role in the cell (Niihama et al., 2009) Attempts to overexpress *CmVPS41* in melon under the 35S promoter were unsuccessful (L. Pascual, unpublished), whereas expression from its own promoter produced fully viable transgenic plants (Giner et al., 2017), suggesting that CmVPS41 protein does not admit neither strong changes in its structure nor in its expression levels. In fact, out of the 12 polymorphisms leading to amino acid changes, only two of them, G85E and L348R, highly conserved residues located in conserved regions (**Supplementary Figure 2**, Giner et al., 2017), have a theoretical strong effect on the protein, and both of them correlate with resistance to CMV-LS. No VPS41 structure has been resolved in any species, and thus, we cannot model the influence of these amino acid changes in the protein structure. However, secondary structure prediction using PROFAcc (https://www.predictprotein. org/) shows that there are two regions in the protein. From amino acid 42 to 399, there is a region with only β sheets, whereas from residue 405 to 825, there are only α helices. Both amino acid 85 and 348 reside in the β sheets region, and both lay in loops between two β sheets (**Figure 3B**). As only subtle changes are allowed in VPS41 protein, it suggests that changes in loops are more easily allowed than others in more structured regions. These loops might lay in the areas involved in the interaction with the viral MP, the determinant of virulence (Guiu-Aragonés et al., 2015), so that these changes would affect only the viral infection without affecting the vital function of transporting cargo proteins to the vacuole. Interestingly, the two parts of the protein nearly correspond to the two parts observed in the gene, since up to nucleotide 1,828, there are more fixed SNPs than from nucleotide 1,829 to the 3′ end of the gene. This way, the first half of the gene, which appears to have admitted more mutations, corresponds to the β-sheet structure and is where the causal mutations reside. The second part of the gene, which is in high linkage disequilibrium, corresponds to the α-helix region. It is possible that this second part of the gene is under stronger selection than the 5′ part.

Melon was likely domesticated twice: in Africa and Asia (Endl et al., 2018) where wild accessions are frequently found. Different *CmVPS41* haplotypes have been identified in wild types from different origins, while most of the analyzed cultivated accessions shared Hap-2. India, and specifically the highly variable momordica group, is considered the origin of most Asian landraces and of the widely commercialized cultivar groups (Blanca et al., 2012; Leida et al., 2015). *CmVPS41* seems to be uniform within the momordica group. The Hap-2, found in all momordica accessions, is also present in ameri, dudaim, and a large set of oriental melons. However, the Far Eastern melons (accessions of the conomon/chinensis/makuwa groups from Japan, China, and Korea), characterized
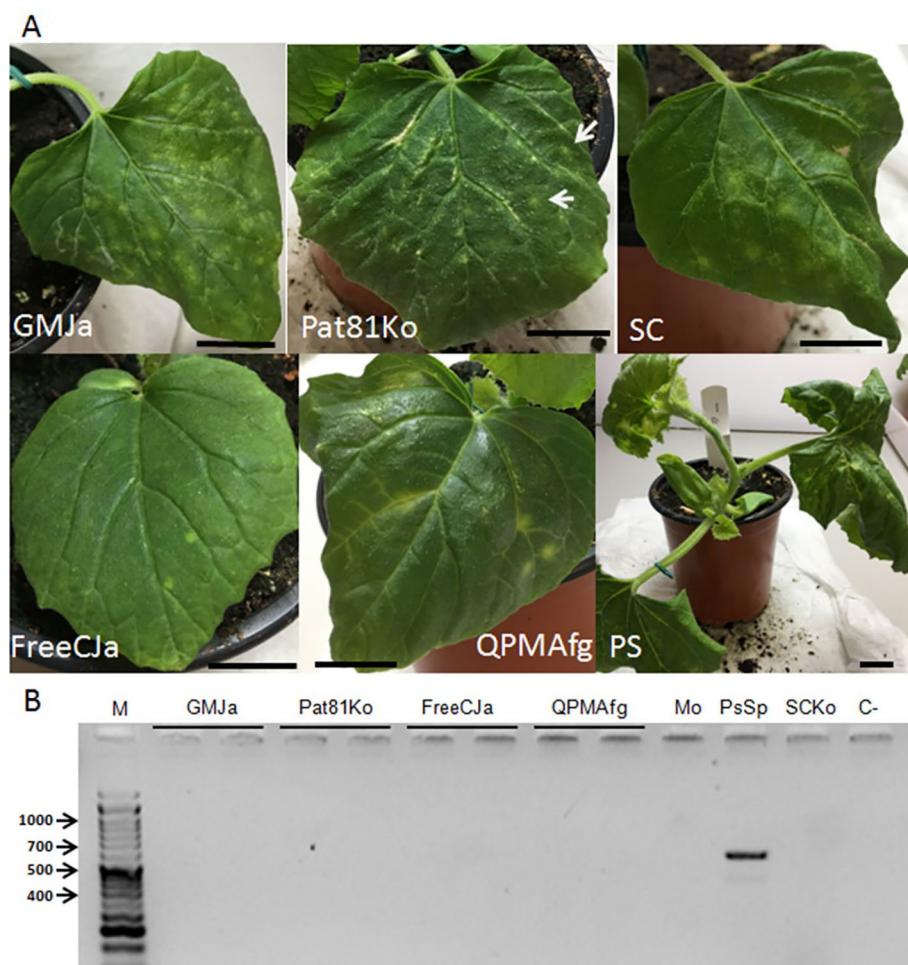
FIGURE 4 | Resistance to CMV-LS in exotic melon genotypes. **(A)** Inoculated leaves from resistant melon genotypes at 11 dpi showing the halos where infection initiates. Black scale bars indicate 2 cm. In PS genotype, at 11 dpi, the whole plant is infected. **(B)** Reverse transcription PCR (RT-PCR) of petiole samples from inoculated leaves of two individual plants per genotype. Mo: Mock inoculated plant. C-: negative control for PCR. M: size marker. On the left, sizes of relevant bands of the marker, in base pairs.

by their low molecular diversity (Blanca et al., 2012; Esteras et al., 2013; Leida et al., 2015), are highly variable for this gene, including accessions carrying three different haplotypes. Two of them, Hap-3 and Hap-4, lead to CMV-LS resistance. These Far Eastern cultivars were originated after the diversification of melon from India towards Far East Asia (Gonzalo et al., 2019); therefore, the mutations related to resistance either could arise in the early diversification of these groups or represent ancestral variability that has been maintained likely by selection. Apart from the previously studied SC, the Japanese cultivars Freeman's Cucumber (Con-FreeCJa), Shiro uri Okayama (Con-ShiroJa), Ginsen makuwa (Con-GMJa), and the Chinese cultivar China51 (Con-Chi51Chi) were previously reported as resistant to CMV (Karchi et al., 1975; Pitrat and Lecoq, 1980; Provvidenti, 1998), although the resistance was not studied in detail. The remaining accessions are new sources of resistance to CMV, and some of them, such as Ogon 9 (Con-OgonChi) and Pat 81 (Con-Pat81Ko), have been previously used in melon

breeding for their resistance to fungus (Perchepied and Pitrat, 2004; Roig et al., 2012).

The Hap-1, which leads to CMV susceptibility, is found in accessions representing not only the main European commercial types, inodorus and cantalupensis, but also in South American and Asian landraces. Therefore, the new resistant accessions sharing the previously characterized SC haplotype and the new ones having the newly described haplotype characterized by a new deleterious mutation are interesting materials for breeding commercial melons against CMV.

Another interesting source of resistance is the dudaim-type Queen's pocket melon, which does not present any putative causal mutation, showing the same protein found in other 25 susceptible accessions. However, Queen's pocket melon is resistant to CMV-LS, and moreover, the resistance is allelic with that of SC as assessed by crossings with the resistant NIL SC12-1-99. Dudaim melons are morphological and molecularly different from other melon groups (Gonzalo et al., 2019). The resistance observed here could be due

to an independent mutation in a noncoding sequence not covered in the current report. For example, it could be due to an alternative splicing which would produce a slightly different protein still able to keep its role in vacuole trafficking. Alternatively, the silent mutation at position 1,263 could have some effect in the RNA folding, with a role in the regulation of the protein. Further work is needed to address if the mutation is located either in a noncoding regulatory sequence of *CmVPS41* or in another tightly linked gene and to understand the events underlying the resistance to CMV-LS in QPM genotype. However, the later possibility is very unlikely, since, given the recessive nature of the resistance, it would mean that, in SC, the other gene would bear also a mutation leading to resistance to CMV-LS, so that, in the F1, the mutation would be in homozygosis. However, in that case, the putative mutation should already be in homozygosis in SC and would have been responsible for the resistance in that genotype. As map-based cloning demonstrated, the only gene responsible for the resistance to CMV-LS in SC was *CmVPS41* (Giner et al., 2017).

VPS41-mediated resistance to CMV in the new resistant genotypes restricts CMV phloem entry, to finally impede the development of a systemic infection. Guiu-Aragonés et al, (2016) reported that, in the resistant introgression line SC12-1-99 (carrying SC resistance gene), the virus movement is blocked in the BS cells that surround the vein, and therefore, it does not reach the phloem to produce a systemic infection. Therefore, our results with the new resistant genotypes again support the role of *CmVPS41* in this resistance. Moreover, the fact that the new resistant accessions belong to four different melon groups suggests that restriction to systemic infection *via* restriction of phloem entry as mechanism for resistance to CMV-LS appeared very early during the melon domestication or diversification. The role of *CmVPS41* in resistance to CMV, and its generality in resistant melon accessions opens new possibilities for using different breeding approaches based on genes that control different steps of the viral cycle, to promote resistances to different viruses. Targeting eIFs genes has provided a useful tool against several viruses in melon (Rodríguez-Hernández et al., 2012; Chandrasekaran et al., 2016) and in other species (Sanfaçon, 2015). A strategy based on producing melon plants pyramiding mutations in eIFs, which would target viral translation/replication, and in VPS41, targeting viral movement, would speed up breeding strategies for sustainable plant protection against viruses of different families.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/**Supplementary Files**.

AUTHOR CONTRIBUTIONS

LP designed and carried out experiments and contributed to write the manuscript. JY did experimental work, MP supervised experimental work, and AM and BP contributed to write the manuscript. AM-H designed experiments and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01219/full#supplementary-material

REFERENCES

Argyris, J. M., Pujol, M., Martín-Hernández, A. M., and Garcia-Mas, J. (2015). Combined use of genetic and genomics resources to understand virus resistance and fruit quality traits in melon. *Physiol. Plant.* 155, 4–11. doi: 10.1111/ppl.12323

Asensio, C. S., Sirkis, D. W., Maas, J. W., Egami, K., To, T.-L., Brodsky, F. M., et al. (2013). Self-assembly of VPS41 promotes sorting required for biogenesis of the regulated secretory pathway. *Dev. Cell* 27, 425–437. doi: 10.1016/j.devcel.2013.10.007

Bandelt, H., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Blanca, J., Esteras, C., Ziarsolo, P., Pérez, D., Fernández-Pedrosa, V., Collado, C., et al. (2012). Transcriptome sequencing for SNP discovery across Cucumis melo. *BMC Genomics* 13, 280. doi: 10.1186/1471-2164-13-280

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17 (7), 1140–1153. doi: 10.1111/mpp.12375

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7, e46688. doi: 10.1371/journal.pone.0046688

Dhillon, N. P. S., Ranjana, R., Singh, K., Eduardo, I., Monforte, A. J., Pitrat, M., et al. (2007). Diversity among landraces of Indian snapmelon (*Cucumis melo*

var. *momordica*). *Genet. Resour. Crop Evol.* 54, 1267–1283. doi: 10.1007/s10722-006-9108-2

Diaz, J. A., Mallor, C., Soria, C., Camero, R., Garzo, E., Fereres, A., et al. (2003). Potential sources of resistance for melon to nonpersistently aphid-borne viruses. *Plant Dis.* 87, 960–964. doi: 10.1094/PDIS.2003.87.8.960

Dogimont, C., Leconte, L., Perin, C., Thabuis, A., Lecoq, H., and Pitrat, M. (2000). Identification of QTLs contributing to resistance to different strains of *Cucumber mosaic cucumovirus* in melon. *Acta Hortic* 510, 391–398. doi: 10.17660/ActaHortic.2000.510.62

Edwardson, J. R., and Christie, R. G., (1991). "Cucumoviruses," in *CRC handbook of viruses infecting legumes*. Eds. J. R. Edwardson and R. G. Christie (Boca Raton, FL: CRC Press), 293–319. doi: 10.1201/9781351071192-16

Endl, J., Achigan-Dako, E. G., Pandey, A. K., Monforte, A. J., Pico, B., and Schaefer, H. (2018). Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. *Am. J. Bot.* 105, 1662–1671. doi: 10.1002/ajb2.1172

Essafi, A., Diaz-Pendon, J. A., Moriones, E., Monforte, A. J., Garcia-Mas, J., and Martin-Hernandez, A. M. (2009). Dissection of the oligogenic resistance to *Cucumber mosaic virus* in the melon accession PI 161375. *Theor. Appl. Genet.* 118, 275–284. doi: 10.1007/s00122-008-0897-x

Esteras, C., Formisano, G., Roig, C., Díaz, A., Blanca, J., Garcia-Mas, J., et al. (2013). SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. *Theor. Appl. Genet.* 126, 1285–1303. doi: 10.1007/s00122-013-2053-5

Fergany, M., Balvir, K., Monforte, A. J., Pitrat, M., Rys, C., Lecoq, H., et al. (2011). Variation in melon (*Cucumis melo*) landraces adapted to the humid tropics of southern India. *Genet. Resour. Crop Evol.* 58, 225–243. doi: 10.1007/s10722-010-9564-6

Giner, A., Pascual, L., Bourgeois, M., Gyetvai, G., Rios, P., Picó, B., et al. (2017). A mutation in the melon vacuolar protein sorting 41 prevents systemic infection of *Cucumber mosaic virus*. *Sci. Rep.* 7, 10471. doi: 10.1038/s41598-017-10783-3

Gonzalo, M. J., Díaz, A., Dhillon, N. P. S., Reddy, U. K., Picó, B., and Monforte, A. J. (2019). Re-evaluation of the role of Indian germplasm as center of melon diversification based on genotyping-by-sequencing analysis. *BMC Genomics* 20, 448–448. doi: 10.1186/s12864-019-5784-0

Guiu-Aragonés, C., Díaz-Pendón, J. A., and Martín-Hernández, A. M. (2015). Four sequence positions of the movement protein of *Cucumber mosaic virus* determine the virulence against *cmv1*-mediated resistance in melon. *Mol. Plant Pathol.* 16, 675–684. doi: 10.1111/mpp.12225

Guiu-Aragonés, C., Monforte, A. J., Saladié, M., Corrêa, R. X., Garcia-Mas, J., and Martín-Hernández, A. M. (2014). The complex resistance to *Cucumber mosaic cucumovirus* (CMV) in the melon accession PI 161375 is governed by one gene and at least two quantitative trait loci. *Mol. Breed.* 34, 351–362. doi: 10.1007/s11032-014-0038-y

Guiu-Aragonés, C., Sánchez-Pina, M. A., Díaz-Pendón, J., Peña, E. J., Heinlein, M., and Martín-Hernández, A. M. (2016). *cmv1* is a gate for *Cucumber mosaic virus* transport from bundle sheath cells to phloem in melon. *Mol. Plant Pathol.* 17, 973–984. doi: 10.1111/mpp.12351

Hashimoto, M., Neriya, Y., Yamaji, Y., and Namba, S. (2016). Recessive resistance to plant viruses: potential resistance genes beyond translation initiation factors. *Front. Microbiol.* 7, 1695–1695. doi: 10.3389/fmicb.2016.01695

Hirai, S., and Amemiya, Y. (1989). Studies on the resistance of melon cultivars to *Cucumber mosaic virus* (I) virus multiplication in leaves or mesophyll protoplasts from a susceptible and a resistant cultivars. *Ann. Phytopath. Soc. Japan* 55, 458–465. doi: 10.3186/jjphytopath.55.458

Karchi, Z., Cohen, S., and Govers, A. (1975). Inheritance of resistance to *Cucumber mosaic virus* in melons. *Phytopathology* 65, 479–481. doi: 10.1094/Phyto-65-479

Kirkbride, J. H. (1993). *Biosystematic monograph of the genus Cucumis (Cucurbitaceae)*. Boone, North Carolina: Parkway Publishers.

Leida, C., Moser, C., Esteras, C., Sulpice, R., Lunn, J. E., de Langen, F., et al. (2015). Variability of candidate genes, genetic structure and association with sugar accumulation and climacteric behavior in a broad germplasm collection of melon (*Cucumis melo* L.). *BMC Genet.* 16, 28. doi: 10.1186/s12863-015-0183-2

Leigh, J. W., and Bryant, D. (2015). Monte Carlo strategies for selecting parameter values in simulation experiments. *Syst. Biol.* 64, 741–751. doi: 10.1093/sysbio/syv030

Malik, A. A., Vashisht, V. K., Singh, K., Sharma, A., Singh, D. K., Singh, H., et al. (2014). Diversity among melon (*Cucumis melo* L.) landraces from the Indo-Gangetic plains of India and their genetic relationship with USA melon cultivars. *Genet. Resour. Crop Evol.* 61, 1189–1208. doi: 10.1007/s10722-014-0101-x

Niihama, M., Takemoto, N., Hashiguchi, Y., Tasaka, M., and Morita, M. T. (2009). ZIP genes encode proteininvolved inmembrane trafficking of the TGN–PVC/vacuoles. *Plant Cell Physiol.* 50, 2057–2068. doi: 10.1093/pcp/pcp137

Perchepied, L., and Pitrat, M. (2004). Polygenic inheritance of partial resistance to *Fusarium oxysporum* f. sp. *melonis* race 1.2 in melon. *Phytopathology* 94 (12), 1331–1336. doi: 10.1094/PHYTO.2004.94.12.1331

Pitrat, M. (2008). "Melon," in *Vegetables I: Asteraceae, Brassicaceae, Chenopodicaceae, and Cucurbitaceae*. Ed. j. a. N., F. Prohens (New York: Springer Science+Business Media, LLC), 283–315.

Pitrat, M. (2017). "Melon genetic resources: phenotypic diversity and horticultural taxonomy," in *Genetics and genomics of Cucurbitaceae.*. Eds. R. Grumet, N. Katzir, and J. Garcia-Mas (Switzerland: Springer International Publisher AG). doi: 10.1007/7397_2016_10

Pitrat, M., and Lecoq, H. (1980). Inheritance of resistance to *Cucumber mosaic virus* transmission by *Aphis gossypii* in *Cucumis melo.*. *Phytopathology* 70, 958–961. doi: 10.1094/Phyto-70-958

Pols, M. S., ten Brink, C., Gosavi, P., Oorschot, V., and Klumperman, J. (2013). The HOPS proteins hVps41 and hVps39 are required for homotypic and heterotypic late endosome fusion. *Traffic* 14, 219–232. doi: 10.1111/tra.12027

Provvidenti, R. (1998). A source of a high level of tolerance to *Squash mosaic virus* in a melon from China. *Cucurbit Genet. Cooperative Rep.* 21, 29–30.

Risser, G., Pitrat, M., and Rode, J. C. (1977). Etude de la résistance du melon (*Cucumis melo* L.) au virus de la mosaïque du concombre. *Ann Amél Plant* 27, 509–522.

Robinson, R. W., and Decker-Walters, D. S., (1997). *Cucurbits*. Wallingford, Oxon, New York: Cab International.

Rodríguez-Hernández, A., Gosalvez, B., Sempere, R., Burgos, L., Aranda, M., and Truniger, V. (2012). Melon RNA interference (RNAi) lines silenced for Cm-eIF4E show broad virus resistance. *Mol. Plant Pathol.* 13, 755–763. doi: 10.1111/j.1364-3703.2012.00785.x

Roig, C., Fita, A., Ríos, G., Hammond, J. P., Nuez, F., and Picó, B. (2012). Root transcriptional responses of two melon genotypes with contrasting resistance to *Monosporascus cannonballus* (Pollack and Uecker) infection. *BMC Genomics* 13, 601–601. doi: 10.1186/1471-2164-13-601

Roossinck, M. J. (2001). *Cucumber mosaic virus*, a model for RNA virus evolution. *Mol. Plant Pathol.* 2, 59–63. doi: 10.1046/j.1364-3703.2001.00058.x

Sabato, D., Esteras, C., Grillo, O., Peña-Chocarro, L., Leida, C., Ucchesu, M., et al. (2019). Molecular and morphological characterisation of the oldest *Cucumis melo* L. seeds found in the Western Mediterranean Basin. *Archaeol. Anthropol. Sci.* 11, 1–22. doi: 10.1007/s12520-017-0560-z

Sanfaçon, H. (2015). Plant translationfactors and virus resistance. *Viruses* 7, 3392–3419. doi: 10.3390/v7072778

Sanseverino, W., Hénaff, E., Vives, C., Pinosio, S., Burgos-Paz, W., Morgante, M., et al. (2015). Transposon insertions, structural variations, and SNPs contribute to the evolution of the melon genome. *Mol. Biol. Evol.* 32, 2760–2774. doi: 10.1093/molbev/msv152

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115–e115. doi: 10.1093/nar/gks596

Fine Mapping of Lycopene Content and Flesh Color Related Gene and Development of Molecular Marker–Assisted Selection for Flesh Color in Watermelon (*Citrullus lanatus*)

Chaonan Wang[1,2], Aohan Qiao[1,2], Xufeng Fang[1,2], Lei Sun[1,2], Peng Gao[1,2], Angela R. Davis[3], Shi Liu[1,2]* and Feishi Luan[1,2]*

[1] *Key Laboratory of Biology and Genetic Improvement of Horticulture Crops (Northeast Region), Ministry of Agriculture, Northeast Agricultural University, Harbin, China,* [2] *College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, China,* [3] *Woodland Research Station, Sakata Seed America, Inc. Woodland, CA, United States*

Lycopene content and flesh color are important traits determined by a network of carotenoid metabolic pathways in watermelon. Based on our previous study of genetic inheritance and initial mapping using F_2 populations of LSW-177 (red flesh) × cream of Saskatchewan (pale yellow flesh), red flesh color was controlled by one recessive gene regulating red and pale yellow pigmentation, and a candidate region related to lycopene content was detected spanning a 392,077-bp region on chromosome 4. To obtain a more precise result for further study, three genetic populations and a natural panel of 81 watermelon accessions with different flesh colors were used in this research. Herein, we narrowed the preliminary mapping region to 41,233 bp with the linkage map generated from F_2 populations of LSW-177 (red flesh) × cream of Saskatchewan (pale yellow flesh) with 1,202 individuals. Two candidate genes, *Cla005011* and *Cla005012*, were found in the fine mapping region; therein *Cla005011* was a key locus annotated as a lycopene β-cyclase gene. Phylogenetic tree analysis showed that *Cla005011* was the closest relative gene in gourd. LSW-177 × PI 186490 (white flesh) and another BC_1 population derived from garden female (red flesh) × PI 186490 were generated to verify the accuracy of the red flesh candidate gene region. By analyzing the expression levels of candidate genes in different developmental stages of different color watermelon varieties, *Cla005011* for the expression differences was not the main reason for the flesh color variation between COS and LSW-177. This indicated that the *LCYB* gene might regulate fruit color changes at the protein level. A new marker-assisted selection system to identify red and yellow flesh colors in watermelon was developed with flesh color–specific CAPS markers and tested in 81 watermelon accessions.

Keywords: lycopene, gene mapping, flesh color, CAPS markers, marker-assisted selection (MAS)

INTRODUCTION

The watermelon (*Citrullus lanatus*) is one of the most important cucurbitaceous crops in the world and occupies approximately 6% of the cultivated area used for all types of vegetables. The main pigment that causes red flesh color in watermelon is lycopene, which is considered one of the most important natural carotenoids in fruits. Lycopene has been a research focus in many areas, including health care products, cosmetics, and nutrition, and has been shown to serve physiological functions in the human body. Lycopene is also the precursor in some physiological, and biochemical processes in plants. Free radical scavenging can damage DNA and proteins, and one effective treatment is high lycopene intake (Mortensen et al., 1997). Cancer prevention, immunity enhancement, and cardiovascular protection can also be improved by lycopene (Feng et al., 2010). Notably, the lycopene in watermelon can be absorbed by the human body directly, but for tomato, another fruit rich in lycopene, dietary lycopene is better absorbed from cooked foods. Lycopene is also the precursor of β-carotene, violaxanthin, and neoxanthin, which participate in different physiological events in plants, including photosynthesis, antenna assembly, and photoprotection (Young, 1991; Latowski et al., 2011).

The visual quality of flesh color in watermelon is an important commodity characteristic for consumers, and different compositions of carotenoids create a wide range of watermelon flesh colors. The main flesh colors for watermelon are white, pale yellow, canary yellow, orange, pink, red, and scarlet. Furthermore, some accessions belonging to the subspecies *Citrullus amarus* are light green. Only trace amounts of phytofluene could be detected in white-fleshed watermelon (Zhao et al., 2008). In watermelons with yellowish flesh, the pigments were found to be a mixture of β-xanthophyll derivatives generated from zeaxanthin, but the composition was different in different accessions. Neoxanthin, violaxanthin, and neochrome were the three main pigments (1.66 and 0.29 µg/g for canary yellow and pale yellow, respectively) (Bang et al., 2010), while all-*trans*-violaxanthin, 9-*cis*-violaxanthin, and luteoxanthin were detected in yellow-fleshed watermelon by Liu (2.01–2.82 µg/g) (Liu et al., 2012). Watermelons with orange flesh had much higher contents of ζ-carotene, prolycopene, or β-carotene than of other pigments (Tadmor et al., 2005). The main pigments found in pink, red, and scarlet watermelon germplasm resources were lycopene, trace amounts of phytoene, prolycopene, and xanthophyll (Perkins-Veazie et al., 2006), while ζ-carotene, zeaxanthin, violaxanthin, and other carotenoids were barely detectable in mature red fruit (Grassi et al., 2013).

The inheritance of red, yellow, and white flesh color in watermelon has been previously reported. Canary yellow (*C*) is dominant to most other flesh colors (*c*), such as red, pink, orange, and pale yellow except white, which was designated *Wf*. The *Wf* gene is epistatic to the yellow flesh trait. Henderson et al. (1998) reported the allele *i-C* (inhibition of *C* and *c*), which was epistatic to *C* and generated red flesh even in the presence of *C*, but these results were not validated by Bang et al. (2010), who studied the genetic basis of red and canary yellow flesh colors in watermelon. In addition, the *py* gene resulting in pale yellow flesh was reported by Bang et al. (2010).

Some major genes or QTLs related to flesh color in cucurbitaceous crops (such as watermelon, melon, and cucumber) have been reported by scientists. In watermelon, the first QTL mapping of flesh color was performed by Hashizume et al. (2003), and two QTLs related to red flesh color were mapped on LG II and VIII with an integrated genetic linkage map. In our previous study, a major QTL (*LCYB 4.1*) conferring lycopene content and red flesh color was located in a 392,077-bp region on chromosome 4 based on a genetic background including red and pale yellow flesh colors (Liu et al., 2015). *ClPHT4;2* were identified to regulate the development of the flesh color through the process of the carotenoid accumulation. In cultivated watermelon varieties, the transcription level of *ClPHT4;2* was controlled by transcription factors *ClbZIP1* and *ClbZIP2*. And the *ClPHT4;2* not only regulate flesh color, but also could control the level of sweetness (Zhang et al., 2016). Subsequently, Branham et al. (2017) had mapped one major gene associated with β-carotene accumulation as regulating the color change of the fruit in watermelon by one yellow fruited line from a cross of NY0016. In melon, a locus associated with β-carotene was detected on chromosome 9 with a high-density linkage map by Harel-Beja et al. (2010). Based on this finding, Tzuri et al. (2015) found that the sequence of the *CmOr* gene in this effective region was homologous to *BoOr* in cauliflower, which controls the accumulation of β-carotene (Lu et al., 2006). Six single-nucleotide polymorphisms (SNPs) differed between the homologous sequences of *CmOr* in orange- and green-fleshed accessions, with only one functional variation (G base to A base) that changed arginine to histidine at the 323rd amino acid position in melon. The expression of *CmOr* was not significantly affected by growth stage in green- and orange-fleshed melons according to reverse transcription–polymerase chain reaction (PCR), implying that the allelic variation in *CmOr* did not affect β-carotene accumulation by altering the transcript level. The same results were also observed in an RNA-seq bulk analysis: most of the genes participating in the carotenoid metabolic pathways exhibited no significant expression changes between the green- and orange-fleshed melon bulks, even though the β-carotene content was significantly different in these two bulks (Chayut et al., 2015). β-Carotene could be detected in callus when the allele of the homologous *CmOr* gene associated with green flesh (arginine into histidine) was transformed into *Arabidopsis* by Tzuri et al. (2015). The F₃ population was obtained by hybridization of white and orange flesh melons, 131 plants were genotyped by RAD-seq, and a major white flesh QTL locus (*CmPPR1*) was found on chromosome 8 (Galpaz et al., 2018). In cucumber, the *ore* gene, which was located on cucumber chromosome 3DS using a RIL (recombinant inbred lines) population, was the key gene for β-carotene accumulation (Bo et al., 2011). Pale yellow flesh color in cucumber was affected by a single recessive gene, *yf*, and fine mapped to a 149.0-kb region on cucumber chromosome 7 containing 22 candidate genes (Lu et al., 2015).

Compared with conventional breeding, rapid breeding through the marker-assisted selection (MAS) method is considered to be an effective strategy for the development of various cultivars. Thus, the investigation of genetic patterns, major gene locations, and specific markers associated with desirable traits should be

top priorities for breeding programs. As relatively few studies of lycopene content and flesh color beyond those cited above have been performed in watermelon, this topic deserves further study. In the present study, we narrowed the lycopene content and flesh color–related QTL region with two genetic populations and investigated the candidate genes. The results provide a research foundation for candidate gene functional validation and support the development of breeding strategies that incorporate traditional and molecular approaches for developing accessions with different flesh colors and MAS in watermelon flesh color breeding.

MATERIALS AND METHODS

Plant Materials

Two red-fleshed lines, "LSW-177" and "garden female" (*C. lanatus* subsp. *vulgaris*); one pale-yellow–fleshed line, "cream of Saskatchewan," abbreviated "COS" (*C. lanatus* subsp. *vulgaris*); and one white-fleshed line, "PI 186490" (*C. lanatus* subsp. *mucosospermus*), were used in our research for genetic population construction. The seeds of LSW-177, COS, and PI 186490 were kindly provided by ARD from the US Department of Agriculture, Agricultural Research Service, South Central Agricultural Research Laboratory. The seeds of garden female were provided by the Laboratory of Molecular Genetic Breeding in Melon and Watermelon, Horticulture College of Northeast Agricultural University, China. Three hybrid combinations segregating for fruit flesh color (LSW-177 × COS, LSW-177 ×

PI 186490, and garden female × PI 186490) were performed to produce the genetic populations. For the LSW-177 × COS and LSW-177 × PI 186490 crosses, the F_1 plants were self-pollinated to produce two F_2 generations (Pop. 1 and Pop. 2) including 352 and 359 individuals, respectively. Meanwhile, we increased population size to 1,202 plants of the same source as Pop. 1 for fine mapping. For the garden female × PI 186490 cross, the F_1 plants were backcrossed to garden female to obtain a BC_1 population (Pop. 3, 222 plants). These three genetic groups were used to describe and validate the inheritance patterns of the flesh color and lycopene content traits in mature fruits. Pop. 1 and Pop. 2 were used in genetic map construction and linkage analysis for gene mining and candidate region reduction; Pop. 2 and Pop.3 were also used to verify the mapping results.

A panel of 81 watermelon accessions (27 accessions were collected from the Germplasm Resources Information Network, US Department of Agriculture, Agricultural Research Service, and others were maintained in the Laboratory of Molecular Genetic Breeding in Melon and Watermelon, Horticulture College of Northeast Agricultural University, China), including white (14 accessions), yellowish (19 accessions), pink (6 accessions), and red (42 accessions) flesh color developed from the natural population, was used for flesh color MAS system construction and verification. All the plants were grown in the greenhouse on the experimental farm of XiangFang and XiangYang Experiment Agricultural Station of Northeast Agricultural University, Harbin (44°04′N, E125°42′), China, during the summer of 2014, 2015, and 2017. The flesh colors of the parental materials and the genetic populations are shown in **Figure 1**.
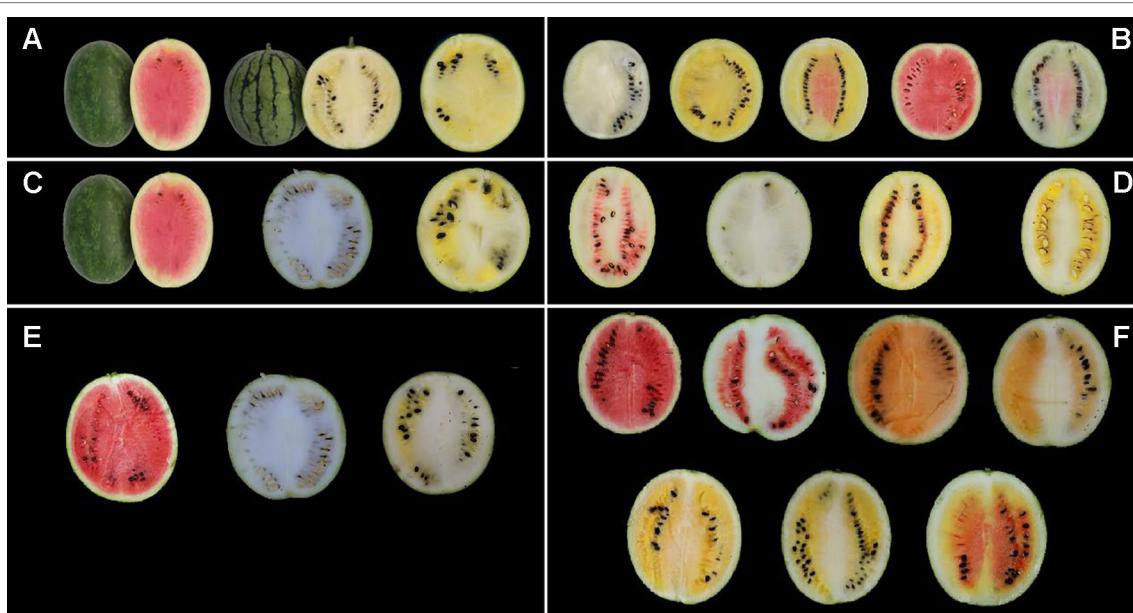
FIGURE 1 | Phenotype pictures of the parental lines, F_1 generation, F_2 populations, and BC_1P_1 population derived from the four parental materials. **(A)** From left to right, LSW-177, COS, and F_1 generation of LSW-177 × COS. **(B)** F_2 generation of LSW-177 × COS segregated into different kinds of flesh color. **(C)** From left to right, LSW-177, PI 186490, and F_1 generation of LSW-177 × PI 186490. **(D)** F_2 generation of LSW-177 × PI 186490 segregated into different kinds of flesh color. **(E)** From left to right, garden female, PI 186490, and F_1 generation of garden female × PI 186490. **(F)** BC_1P_1 generation of garden female × PI 186490 segregated into different kinds of flesh color.

Lycopene Measurement and Flesh Color Evaluation

All the plants were artificially pollinated with an identification tag showing the date of pollination, and the fully mature fruits were harvested between 35 and 50 days after pollination. The harvested fruits were cut longitudinally, photographed for reassessment, and categorized into different flesh groups by visual observation. Flesh color data were scored as 3 for red, 2 for canary yellow, and 1 for pale yellow in Pop. 1, whereas 1 was used for red and 0 for nonred in Pop. 2 and Pop. 3. Pulp samples were mixed equally with five parts taken from the central tissue and four other points around the edge of the mature fruits. The samples were maintained at −80°C for high-performance liquid chromatography (HPLC) analysis. A Waters PDA detector 2535 and Agilent LC ZORBAX SB-C18 column (4.6 × 250 mm, 5 μm) were used for lycopene content analysis. Lycopene authentic standard (Sigma, USA) was dissolved with dichloromethane to obtain a 5 μg/mL concentration solution. The retention time was referenced to identify lycopene with the UV-V spectra condition of 472 nm. Data were analyzed by BREEZE software (Waters), and the lycopene content was quantified by utilizing authentic standard curves and determined as μg/g flesh weight. The method and protocol of lycopene extraction and measurement were according to Yuan et al. (2012) and Liu et al. (2015).

Statistical Analysis

The mean values, standard deviations, trait distributions, and χ^2 test analysis, and other statistical analyses were evaluated with SPSS 19.0 software (SPSS Inc., United States)

DNA Preparation

Young leaf tissues of 15 plants from each parental material and F_1 generation were used for DNA extraction. The DNA of Pop. 1, Pop. 2, Pop. 3, fine mapping population plants, and the 81 accessions was extracted individually with a modified CTAB (hexadecyl trimethyl ammonium bromide) method as reported previously (Luan et al., 2008).

Marker Development, PCR, and Gel Electrophoresis
Caps Markers

CAPS markers were developed based on resequencing of the four parental watermelon lines with the Illumina HiSeq 2000 high-throughput sequencing platform. In total, 10 Gb of data were obtained from each material, covering the watermelon genome at more than 20×. The resequencing data analysis and marker development were performed according to Liu et al. (2015).

Twenty-two restriction endonucleases (*Eco*RI, *Bsa*HI, *Hind*III, *Mbo*II, *Pst*I, *Sca*I, *Bam*HI, *Mlu*I, *Alu*I, *Dra*I, *Pvu*I, *Kpn*I, *Ras*I, *Hind*II, *Taq*I, *Msp*I, *Sac*I, *Bcl*I, *Xho*I, *Ned*I, *Sac*I, and *Mbo*I) (Thermo Scientific, Massachusetts, United States) were used for marker development. CAPS loci were selected

on each chromosome for primer design, and amplicons derived from the DNA of both parents and the F_1 generation were digested with restriction endonucleases. The PCR mixture and conditions for the CAPS markers were reported previously (Liu et al., 2015). The reaction mixture and conditions for the enzyme digestions were determined from the manual of each restriction endonuclease (Thermo Scientific). Finally, 1% agarose gel electrophoresis was used to examine the enzyme-digested products.

SSR and Indel Markers

In total, 449 pairs of watermelon SSR markers (including 23 pairs of core watermelon SSR markers, Zhang et al., 2011) and 556 pairs of melon SSR markers were used for polymorphism selection. All SSR markers were derived from available publications (Danin-Poleg et al., 2001; Fazio et al., 2002; Silberstein et al., 2003; Yi et al., 2003; Gonzalo et al., 2005; Joobeur et al., 2006; Zalapa et al., 2007; Fernandez-Silva et al., 2008). Eighteen pairs of SSR and indel markers (including 17 SSR markers and 5 indel markers) used for mapping yellow fruit flesh traits in cucumber (Lu et al., 2015) were also chosen in this research. The PCR mixtures for SSR and indel amplification were the same as those used for the CAPS markers, and the conditions were according to Lu et al. (2015). The PCR products were analyzed with 6% denaturing polyacrylamide gel electrophoresis and visualized by silver staining.

Map Construction and Secondary and Fine Mapping of the Candidate Region

Cleaved amplified polymorphism sequence (CAPS) markers were developed both in the preliminary mapping region and throughout the whole genome for fine mapping. Twenty-six pairs of new CAPS primers were designed in the preliminary mapping region of *LCYB 4.1*, among which eight pairs showed polymorphisms between LSW-177 and COS. Some CAPS markers were also designed based on a sequence alignment with zeaxanthin epoxidase, which has been reported to show differential expression among red-, yellow-, and white- fleshed watermelon during growth period (Nakkanong et al., 2012 and Lv et al., 2015). At the same time, the SSR marker *SSR17292* (Lu et al., 2015) for identifying yellow flesh in cucumber performed polymorphism between LSW-177 and COS was used. Subsequently 1,202 extended population (from the same source of Pop. 1) were used for fine mapping. And in the candidate region, 15 pairs of CAPS were developed in the fine mapping interval, and 4 pairs of polymorphic markers were included.

Genetic linkage map construction was performed using the IciMapping V3.3 software (Institute of Crop Science Chinese Academy of Agricultural Sciences, Beijing, China; Meng et al., 2015), and all the markers were grouped at a minimum logrithm of odd (LOD) score of 6.0. The software package Map Chart 2.1 (Plant Research International, Wageningen, the Netherlands; Voorrips, 2002) was used to graphically represent the linkage groups in the map. QTL analysis was also performed with IciMapping V3.3 (Meng et al., 2015), and

QTLs with a LOD score ≥5.0 were considered as the available loci for detection. The fine mapping region was narrowed with the recombinant plants.

Sequence Annotation and Gene Prediction in the Genomic Region Harboring the Target Gene

The genome data of the four parental materials and another 20 watermelon accessions that have been published (Guo et al., 2012) were used for sequence alignment with the watermelon reference genome sequence (97103) from the Cucurbit Genomics Database (Zheng et al., 2018). Using the open reading frame (ORF) and Basic Local Alignment Search Tool (BLAST; Altschul et al., 1900) analysis to detect the candidate genes in the mapping region, only sequences that matched with an identity of more than 95% were retained. The sequence of the coding sequence (CDS) and candidate genes were aligned to detect the splice sites using the Splign tool (Kapustin et al., 2008). Candidate gene sequence alignment and amino acid variation were performed using DNAMAN 6.0 (Lynnon Biosoft, USA) software.

Validation of Watermelon Flesh Color Markers for MAS Breeding

To develop practical molecular markers for MAS breeding, the markers cosegregating with the lycopene content and flesh color locus were validated using Pop. 3 and 81 watermelon accessions to investigate the correlation between the genotype and phenotype data.

Construction of Phylogenetic Tree

In the Cucurbitaceae genome database, the *LCYB* and *LCYE* homologous genes of different Cucurbitaceae crops were searched, with the addition of an *LCYB* gene from tomato, and then the encoded amino acid sequences were obtained. The amino acid sequences were compared by multiple sequence alignment using trimAI software (Capella-Gutierrez et al., 2009), and then MEGA (Kumar et al., 2018) was applied. A phylogenetic tree was established with the UPGMA function in MEGA X (Kumar et al., 2018) software.

Transcript Expression Level of Candidate Genes in Different Watermelon Varieties

In order to explore the expression level of candidate genes in different flesh color watermelons, species 97103, PI 1296341, COS, and LSW-177 were selected for the goal varieties from the database http://cucurbitgenomics.org/ (the RNA-sseq data of COS and LSW-177 were the target varieties in this research). Based on the transcript expression (RPKM) of candidate genes *Cla005011* and *Cla005012* in the above four varieties, we aggregated their data and produced a trend graph using software GraphPad 8.0. We analyzed the expression levels of the two candidate genes at different mature stages of different flesh color watermelons.

RESULTS

Phenotypic Segregation Analysis of Flesh Color and Lycopene Content

The flesh color of the F_1 generation was canary yellow in Pop. 1, close to pale yellow, suggesting pale yellow with an incomplete dominance over red. Five categories of flesh color, red (87 plants), pale yellow (48 plants), canary yellow (173 plants), and two irregular color patterns consisting of red mixed with pale and canary yellow or red in mixed patterns in the heart and placental tissues of the fruit (18 and 26 plants, respectively), were found in the segregating population. Most of the mixed pale and canary yellow fruits had flesh color >50% canary or pale yellow by cross-sectional area, so the two mixed-color plants could be classified as canary yellow and pale yellow. According to these classification criteria, 199 (173 + 26), 66 (48 + 18), and 87 plants were judged to have canary yellow, pale yellow and red flesh color in Pop. 1, fitting a genetic segregation ratio of 9:3:4 (χ^2 = 0.02 and 1.12, P = 0.99 and 0.57 for the year of 2013 and 2014, respectively), which indicated that flesh color was affected by two major genes. Canary yellow and pale yellow could also be classified into red and nonred groups by visual observation. The segregation of these two groups yielded a ratio that did not differ significantly from a 3:1 ratio (nonred group: red group = 265:87, $P < 0.05$) by statistical analysis. These results indicated that a single major recessive gene determined red and nonred color in watermelon based on the genetic background of Pop. 1.

High-performance liquid chromatography analysis of lycopene contents in mature fruit showed that LSW-177 had a high lycopene content with an average value of 41.72 ± 2.82 μg/g, much higher than 0.24 ± 0.03 and 0.42 ± 0.05 μg/g measured in COS and the F_1 generation, respectively. Comparing the lycopene content and flesh color data, we observed that the plants showed a red flesh color when the lycopene content was higher than 13.57 μg/g. Pop. 1 could also be divided into two groups (high- and low-lycopene groups) at this threshold value. The genetic ratio of the high-lycopene (87 plants) to low-lycopene (265 plants) groups in the F_2 progeny adequately fitted a 1:3 (χ^2 = 0.015, P = 0.902) ratio, demonstrating that one major gene affects lycopene accumulation. We also compared these results with our previously published data (Liu et al., 2015) from 2013, when we collected lycopene content data in an F_2 generation (234 plants) with the same parental materials. The F_2 generations in both years showed the same genetic ratio for flesh color and similarly divergent trends in lycopene content. The results of the χ^2 goodness-of-fit test of the segregation ratios and the lycopene content separation analysis in the F_2 populations derived from LSW-177 and COS are shown in **Table 1** and **Figure 2**.

In Pop. 2, the flesh color of the F_1 generation was canary yellow mixed with white. Four main flesh colors, red mixed with white, red mixed with white and yellow, canary yellow mixed with white, and white, emerged in the F_2 generation. For the red and white mixed groups, most of the plants accumulated lycopene only around the seed region, while only two plants had fully red flesh color. Two groups were separated based on flesh color, one group with red or red mixed with white flesh and the

TABLE 1 | Flesh color separation proportion of the parental materials and F_2 population.

Generation	Canary yellow	Pale yellow	Red	Expected ratio	χ^2	P
P_1	0	0	30	0:0:1	0	1
P_2	0	25	0	0:1:0	0	1
F_1	30	0	0	1:0:0	0	1
F_2 (2014)	199	66	87	9:3:4	0.02	0.99
F_2 (2013)	129	50	55	9:3:4	1.12	0.57
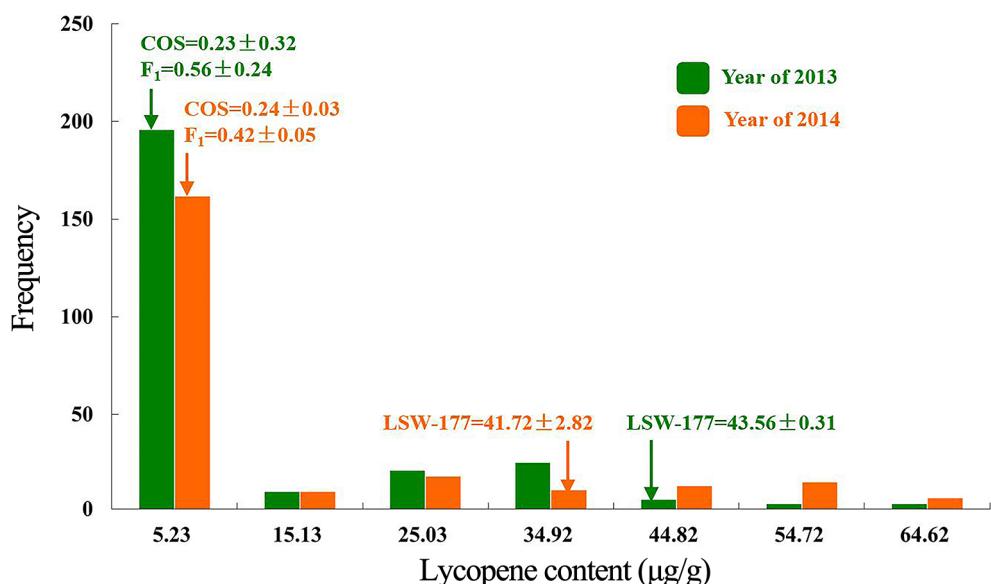
FIGURE 2 | Segregation analysis of lycopene content in the F_2 population from a cross between LSW-177 and COS parental strains in the year of 2013 and 2014.

other group with yellow, white, and mixed yellow and white flesh. The numbers of individuals in these two groups were 97 and 262, fitting a genetic segregation ratio of 1:3 ($\chi^2 = 0.780$, $P = 0.377$).

In Pop. 3, the F_1 plants had canary yellow mixed with white flesh. Approximately seven main flesh colors (red, red mixed with white, orange, orange mixed with white, canary yellow, canary yellow mixed with white and mixed canary yellow, orange and red) segregated in the BC_1P_1 generation. According to the classification standards of Pop. 2, two groups were also separated, one with red and red mixed with white, and the remaining fruits with other flesh colors. The genetic segregation of these two groups was 116:106, fitting a ratio of 1:1 ($\chi^2 = 0.450$, $P = 0.502$). Considering that the F_1 plants of Pop. 2 and Pop. 3 did not have white flesh color similar to that of PI 186490, the white flesh trait has incomplete dominance over red flesh in watermelon. The separate proportions of the three genetic populations are shown in **Table 2**.

Secondary Mapping of the Lycopene Content and Red Flesh Color Loci Using Genome Resequencing Data

Two linkage maps consisting of 311 markers (274 CAPS, 37 SSR) and 200 CAPS markers were constructed from Pop. 1 and Pop. 2 (**Figures S1** and **S2**). Eight new CAPS markers were developed in the initial mapping region for secondary mapping (**Figure 3**). The order of the CAPS marker locations did not tightly correspond to the reference genome. For Pop. 1, one major QTL that related to both the lycopene content and red flesh color traits shared the same candidate region on chromosome 4 between the newly developed CAPS markers *WII04E08-38* and *WII04EBsaHI-6*, 0.15 and 0.05 cM away from each marker, with high R^2 values (84.5% and 81.5%) and LOD scores of 86.3 and 91.21, respectively. Based on the genome resequencing data, the physical distance between *WII04E08-38* and *WII04EBsaHI-6* was 92,931 bp. The 11 CAPS

TABLE 2 | The separate proportion of red and nonred plants in the three genetic populations.

Population	Red	Nonred	Expected ratio	χ^2	P
Pop. 1	87	265	1:3	0.015	0.902
Pop. 2	97	262	1:3	0.780	0.377
Pop. 3	116	106	1:1	0.450	0.502

FIGURE 3 | Genetic linkage groups of chromosome 4 for QTL detection and candidate genes in the effective region. **(A)** Genetic linkage group of chromosome 4 with the F$_2$ generation (359 plants) derived from a cross between garden female and PI 186490. **(B)** Genetic linkage group of chromosome 4 with the F$_2$ generation (352 plants) derived from a cross between LSW-177 and COS. **(C)** Linkage analysis of the CAPS markers in the QTL effective region. **(D)** Candidate genes in the QTL detected region.

markers (from *WII04E07-40* to *CAPSSacI* on chromosome 4; **Figure 3B**) confirmed the chromosome region, with 12 candidate genes cosegregating with flesh color in most individuals of Pop. 1. All the mixed-color plants were heterozygous, similar to the F$_1$ generation.

Five candidate genes (*Cla005011* to *Cla005015*, which are presumed to encode proteins, **Table S1**) were detected in the 92,931-bp region (*WII04E08-38* to *WII04EBsaHI-6*) by consulting the Cucurbit Genomics Database (Zheng et al., 2018). Based on the results of ORF and BLAST (Altschul et al., 1900) analysis, the lycopene β-cyclase (*LCYB*) mRNA, LCYB-red allele, and CDS were located in this region with high sequence similarity to *Cla005011* (8,886,138 to 8,887,652), which indicated that although *SSR17292* related with yellow flesh color in cucumber (Lu et al., 2015) and also exhibited polymorphic between LSW-177 and COS, no QTL locus was detected. In Pop. 2, a major QTL (with an R^2 of 70.54%) between LSW-177 and PI 186490 was preliminarily located in the same region as in Pop. 1, with the flanking CAPS markers of *WIII4-440* (8,348,876) and *WIII4-435* (10,211,848) on chromosome 4. We also used LSW-177 and PI 186490 to determine the polymorphism status of the 11 CAPS markers that cosegregated with flesh color in Pop. 1, and four markers (*WII04EBsaHI-6*, *WII04E08-38*, *CAPSMboI*, and *CAPSHindII*) were found to be

polymorphic. Based on the results of the linkage analysis, only *WII04EBsaHI-6* was assigned to the region between *WIII4-440* and *WIII4-435*, and it narrowed the effective QTL region with an R^2 value of 64.60%. *CAPSMboI* was polymorphic between LSW-177 and PI 186490, but in Pop. 2, all the individuals had the same alleles as LSW-177. The genotyping data showed that *WII04E08-38* and *CAPSHindII* had distorted segregation in Pop. 2, but these markers were distant from *WII04EBsaHI-6* and the target region. This result may reflect the fact that PI 186490 belongs to *C. lanatus* subsp. *mucosospermus*, which is another subspecies of watermelon. Considering the genome divergence of different watermelons, separate reference genomes are needed. For the CAPS marker *WII04EBsaHI-6*, all plants having red mixed with white flesh color showed the same electrophoretic bands as LSW-177, and the individuals with mixed flesh colors were heterozygous, similar to the F$_1$ generation.

Fine Mapping of the Red Flesh Color Loci

By filtering the recombinant plants using CAPS markers *WII04E08-33* and *WII04E08-40* in the F$_2$ population, which was derived from the enlarged population of Pop. 1 containing 1,202 individuals, 28 recombinant individuals were found. Among these,

4 dominant homozygous individuals, 6 recessive homozygous individuals, and 18 recombinant individuals were identified. Four CAPS markers (CAPSNedI-1, CAPSEcoRI4-8, CAPSRsaI4-21, CAPSSacI4-14) and one known marker, WII04EBsaHI-6, were developed between the WII04E08-33 and WII04E08-40 markers. The target region between markers WII04EBsaHI-6 and CAPSRsaI4-21 was ultimately acquired by using the above markers to fine map the six recessive homozygous individuals (red flesh), and the physical distance was 41.2 kb (Figure 4). Using the Cucurbit Genomics Database (Zheng et al., 2018), two candidate genes, Cla005011 and Cla005012, were found in the fine mapping region between the positions from 8,886,138 to 8,926,873.

Upon comparison and analysis of the candidate gene sequences between the two parents with the resequencing data, three SNP mutations were found in the exon of the candidate gene (Cla005011), among which two SNPs led to two amino acid changes (mutation of the 676th base G/C to T/A resulted in the 226th amino acid changing from valine [V] to phenylalanine [F], and mutation of the 1,234th base G/C to C/G resulted in the 435th amino acid changing from lysine [K] to asparagine [N]). No amino acid changes were detected with the third SNP mutation (although the 12th base was mutated from A/T to G/C, it did not cause amino acid changes; Figure 5). A single-nucleotide mutation within the exon of the candidate gene (Cla005012)

led to one amino acid change (the 821th base had a G/C to A/T mutation, resulting in the 274th amino acid changing from arginine [R] to histidine [H]; Figure 6). Furthermore, according to fine mapping and combination of the results of CDS blasting with the gene LCYB, Cla005011 was speculated to be on the same locus as LCYB in watermelon.

Genetic Effect of LCYB4.1 on Lycopene Content in Pop. 1

According to the genotyping data, the 11 CAPS markers perfectly cosegregated with flesh color and lycopene content in Pop. 1. To further validate the results of the QTL mapping, we also tested the relationships between lycopene content and the CAPS marker allele pairs (WII04EBsaHI-6 and WII04E08-38) for all individuals in Pop. 1 (Table S2). For the CAPS marker WII04EBsaHI-6, the average content of lycopene in plants with homozygous LSW-177 alleles was $37.38 \pm 13.24 \, \mu g \cdot g^{-1}$, while the plants with homozygous COS alleles had a value of $0.23 \pm 0.06 \, \mu g \cdot g^{-1}$. Heterozygous individuals accumulated an average content of $2.62 \pm 8.10 \, \mu g \cdot g^{-1}$. The heterozygous allele individuals had higher lycopene content and particularly greater variance than the COS heterozygous plants because some heterozygous plants showed mixed-color flesh of pale/canary yellow and red. The same trend was also observed for WII04E08-38.
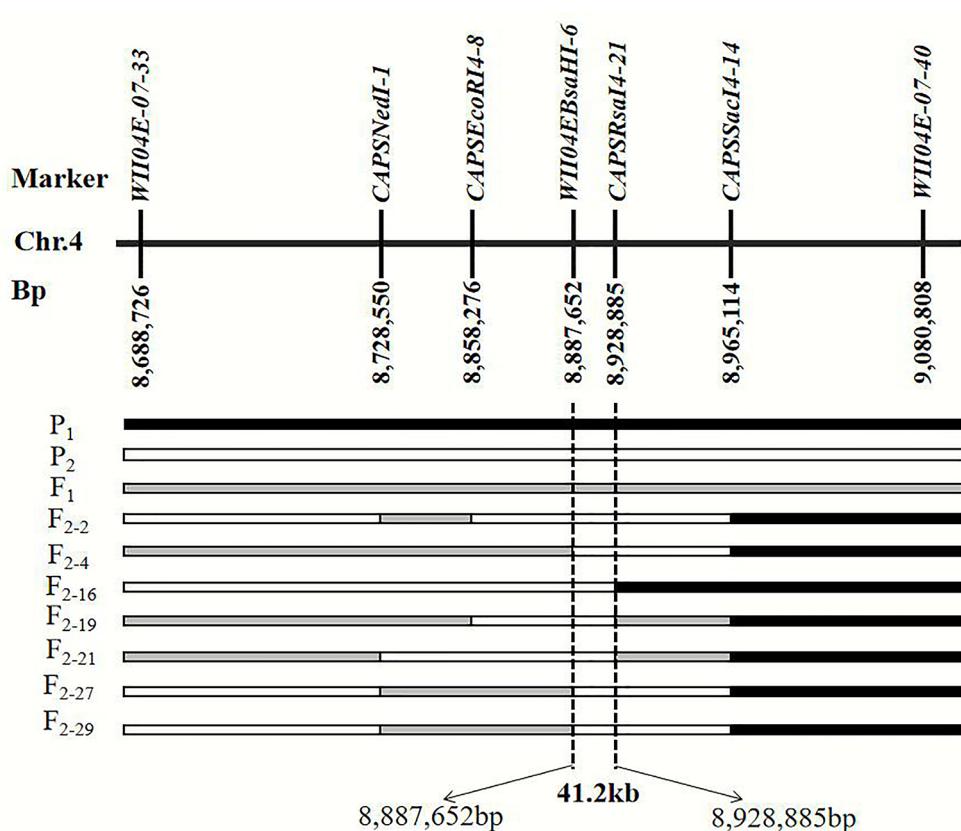
FIGURE 4 | Fine mapping of the genes that control the color of watermelon red flesh through recombinant recessive individuals. P₁ black frame represents the parent material LSW-177, P₂ white frame represents the parent material PI 186490, F₁ was gray frame. F₂₋₂, F₂₋₄, F₂₋₁₆, F₂₋₁₉, F₂₋₂₁, F₂₋₂₁, F₂₋₂₇, and F₂₋₂₉ were recessive recombinant individuals, respectively.
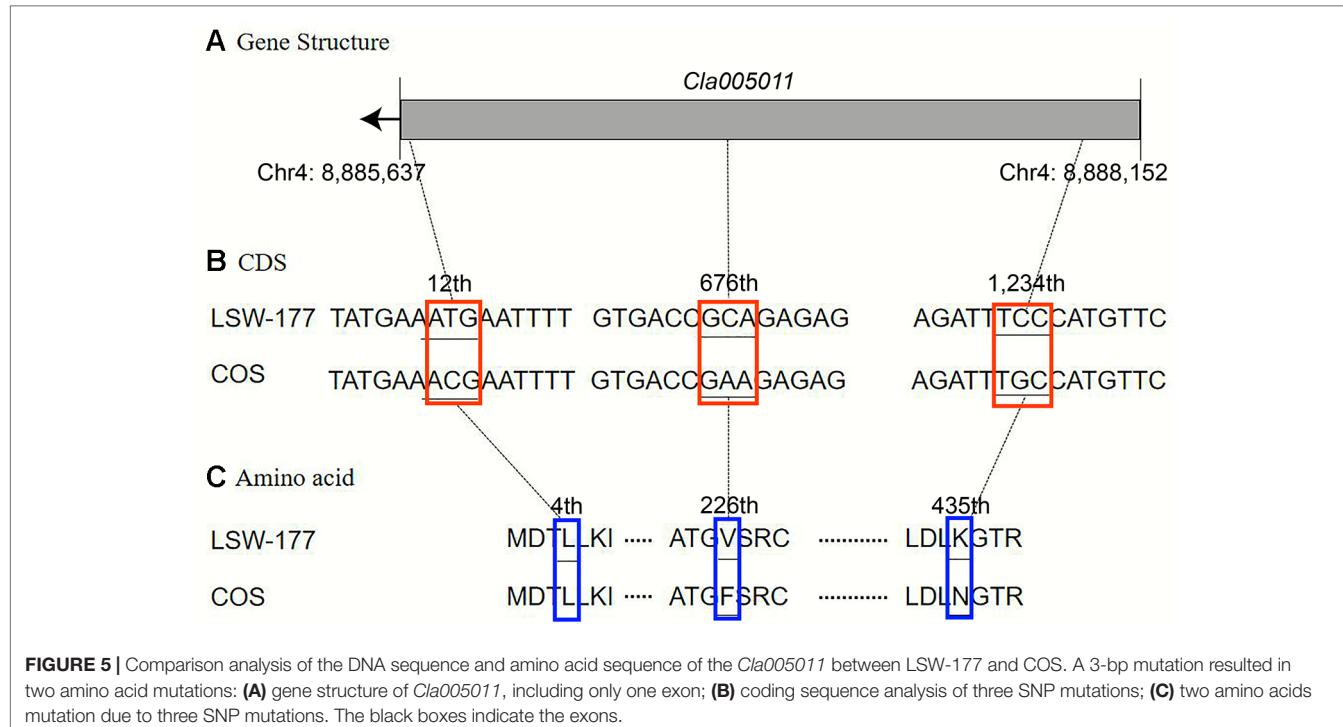
FIGURE 5 | Comparison analysis of the DNA sequence and amino acid sequence of the *Cla005011* between LSW-177 and COS. A 3-bp mutation resulted in two amino acid mutations: **(A)** gene structure of *Cla005011*, including only one exon; **(B)** coding sequence analysis of three SNP mutations; **(C)** two amino acids mutation due to three SNP mutations. The black boxes indicate the exons.
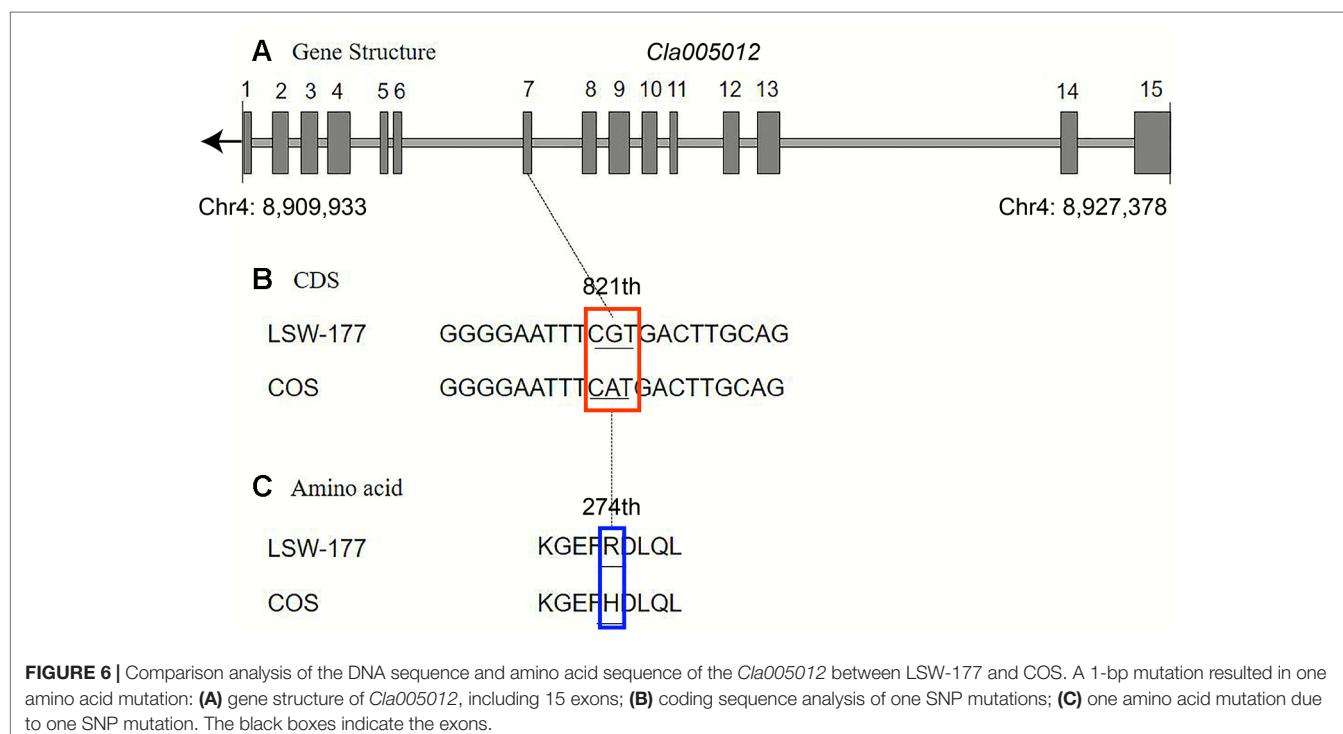
FIGURE 6 | Comparison analysis of the DNA sequence and amino acid sequence of the *Cla005012* between LSW-177 and COS. A 1-bp mutation resulted in one amino acid mutation: **(A)** gene structure of *Cla005012*, including 15 exons; **(B)** coding sequence analysis of one SNP mutations; **(C)** one amino acid mutation due to one SNP mutation. The black boxes indicate the exons.

Comparison of LCYB Sequences in Watermelon

The genome data of 4 parental materials and 20 other published watermelon accessions (Guo et al., 2012) were used to extract the sequence of the *LCYB* gene for sequence alignment and variation detection. The 24 watermelon samples exhibited four flesh colors: red, yellowish, green, and white. According to the results, a total of 20 SNP loci were detected in the *LCYB* sequence among these 24 watermelons, with only one SNP distinguishing red-fleshed color (G^{676th}) plants from nonred

(T^{676th}) accessions, except in PI 248178, which had white flesh but also had a G base at the 676th position in the exon region. This SNP (G^{676th} to T^{676th}) changed valine (V: red) into glycine (nonred). It was also the position of the restriction site of the CAPS marker *WII04EBsaHI-6* (**Figure S3**).

Marker Utilization and MAS for Flesh Color in Watermelon

To verify the applicability of these markers in different genetic populations and watermelon accessions, the 11 CAPS markers in the effective region were used for genotyping in Pop. 3 and a watermelon panel. Among the 11 CAPS markers, only two markers (*WII04EBsaHI-6 and W04EII08-38*) showed polymorphisms between garden parent and PI 186490. In Pop. 3, two flesh colors (red and red mixed with white) appeared in individuals homozygous for the same alleles as garden parent, while plants with other flesh colors carried heterozygous alleles, similar to the F_1 generation.

Considering the flesh color–specific markers and the different polymorphisms of white-fleshed plants, *WII04EBsaHI-6*, *W04EII08-38*, *WII04EKpnI-1*, and *WII04E07-40* were selected to construct a MAS approach for red, yellowish, and white flesh color in 81 watermelon accessions. Based on the genotyping results, the CAPS markers *WII04EBsaHI-6* and *W04EII08-38* cosegregated with red and pink flesh color in these accessions, except in two PI lines (PI 193963 and PI 601228, which had yellowish flesh but exhibited the same restriction enzyme fragments as the red and pink plants). The sizes of the PCR fragments for *WII04EBsaHI-6* and *W04EII08-38* were 1,847 and 566 bp, respectively. Plants with pink and red flesh colors showed enzyme-cut fragments of 1,182 and 665 bp for *WII04EBsaHI-6* and 566 bp for *W04EII08-38*, while the yellowish- and white-fleshed accessions showed fragments of 1,847 and 287 or 279 bp for these two markers (since the resolution ratio of the 1% agarose gel was 100 bp, the 287- and 279-bp products appeared as one band) (**Figure S4**).

For the CAPS markers *WII04EKPnI-1* and *WII04E07-40*, the PCR amplicons were 517 and 618 bp for all 81 accessions. In the yellowish group, each plant exhibited 389- or 128-bp bands for *WII04EKPnI-1* and 460- or 158-bp bands for *WII04E07-40*. The pink, red, and white plants showed 517- and 618-bp bands with these two markers. Three white-fleshed PI lines (PI 248774, PI 532666, and PI 254622) had the same results as the yellowish-fleshed plants. Some white-fleshed PI lines were heterozygous at these marker loci. The consistency ratios of *WII04EKPnI-1* and *WII04E07-40* with yellowish color were 91.4% and 92.6%, respectively (**Figure S5**).

To distinguish the three types of flesh color (red, yellowish, and white), the four markers were combined for further testing. Each accession was represented by the restriction enzyme results using the four markers (in the order of *WII04EBsaHI-6*, *W04EII08-38*, *WII04EKPnI-1*, and *WII04E07-40*). Plants with the marker restriction site were scored as 1, those without were scored as 0, and heterozygotes were recorded as h. With this method, each material could be given a marker code for flesh color. The results showed that the combination of these four markers could clearly distinguish the three flesh colors. According to the genotyping results (**Table S3**), all of the red and pink plants had the code 1, 0, 0, 0, whereas the code for the yellowish plants was 0, 1, 1, 1. For white-fleshed plants, the codes were not uniform, and heterozygous loci were present in some plants, but the code of each white-fleshed watermelon was quite different from those of the pink, red, and yellowish accessions. With these marker codes, we can easily distinguish the three types of flesh color in watermelon. The SSR marker *SSR17292* also showed different polymorphisms in the 81 watermelon accessions but did not cosegregate with flesh color.

Expression Levels of Candidate Genes in Different Watermelon Varieties Based on Published RNA-Seq Data

To explore the expression levels of select candidate genes in watermelon varieties with different flesh colors, four watermelon varieties were selected (according to the website http://cucurbitgenomics.org/): species 97103 (red flesh), PI 296341 (white flesh), COS (pale yellow flesh), and LSW-177 (red flesh). The transcriptional expression (RPKM) of two candidate genes, *Cla005011* and *Cla005012*, in the above species of watermelons was quantified at different periods of development. In all the accessions, COS, LSW-177, 97103 (red flesh), and PI 296341 (white flesh), the expression of *Cla005011* did not explain the lycopene accumulation difference between different watermelon accessions. *Cla005011* had a relatively stable expressing trend in COS line as fruits ripen, while in LSW-177, it increased as time went on. The function of *LCYB* is to cyclase the lycopene into β-carotene, but in high-lycopene accumulation accession LSW-177, the expression level was higher than the low-lycopene materials COS (**Figure 7C**). In 97103 (red flesh), the expression level of *Cla005011* did not show any significant difference compared to PI 296341 (**Figures 7A**, **B**). *Cla005012* was annotated as the kinesin-like protein that seemed to be not related with lycopene accumulation. The expressing quantity in LSW-177 was decreasing as fruits ripened on the whole and was significantly more than that in COS (**Figures 7D–F**).

Phylogenetic Tree Analysis of the Lycopene Cyclase Gene Family

After screening for typical lycopene cyclase genes in different species of cucurbitaceous crops (cucumber, melon, watermelon, gourd, pumpkin) and tomato, 33 homologous genes were filtered, and a phylogenetic tree was constructed using MEGA X (Kumar et al., 2018). The results showed that all genes were divided into three categories, which were defined as lycopene β-cyclase (*LCYB*) and lycopene epsilon-cyclase (*LCYE*). *LCYB* clusters I and II contained 13 and 9 genes, respectively. Additionally, the number of exons was relatively consistent and conservative. Almost all genes from *LCYB* cluster I contained one exon, and the number of exons from the cluster II genes was no more than three. The remaining 11 genes from cluster III encoded *LYCE*, which
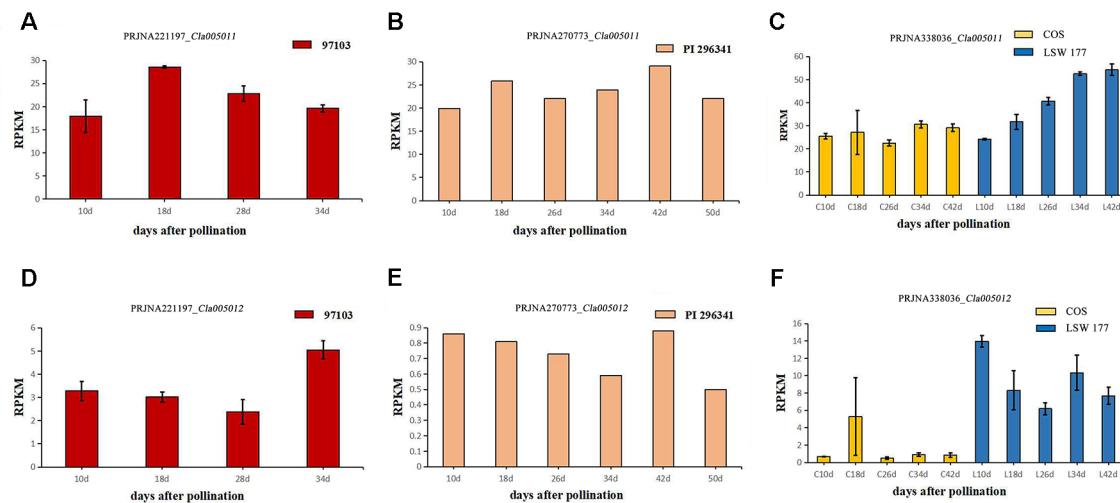
FIGURE 7 | The expression levels of two candidate genes *Cla005011* and *Cla005012* in four different watermelon varieties (97103, LSW-177, COS, and PI 296341) at different maturity stages of the flesh. *Cla005011*: **(A)** BioProject: *PRJNA221197* (RPKM of 97103), sample DAP10, DAP18, DAP28, DAP34, DAP: days after pollination. **(B)** BioProject: *PRJNA270773* (RPKM of PI 296341), sample DAP10, DAP18, DAP26, DAP34, DAP42, DAP50. **(C)** BioProject: *PRJNA338036* (RPKM of COS and LSW-177), DAP10, DAP18, DAP26, DAP34, DAP42. *Cla005012*. **(D)** BioProject: *PRJNA221197* (RPKM of 97103), sample DAP10, DAP18, DAP28, DAP34. **(E)** BioProject: *PRJNA270773* (RPKM of PI 296341), sample DAP10, DAP18, DAP26, DAP34, DAP42, DAP50. **(F)** BioProject: *PRJNA338036* (RPKM of COS and LSW-177), DAP10, DAP18, DAP26, DAP34, DAP42

regulates lycopene to generate the synthesis of α-carotenoids, which in turn produces lutein. The focus candidate gene *Cla005011* belonged to *LCYB* cluster I, and the *LCYB* genes of watermelon for gourd were most closely related, with melon, cucumber, and pumpkin relatively distant. Tomato species with higher levels of lycopene accumulation were used in the process of constructing the phylogenetic tree, and the *LCYB* genes of watermelon were further related to that of tomato. The phylogenetic tree analysis demonstrates that the candidate gene identified from the fine mapping results participated in regulating red flesh formation through the process of carotenoid production (**Figure 8**).

DISCUSSION

Inheritance of Flesh Color and Lycopene Content

Red flesh color and lycopene content are major influences on watermelon quality and consumption. In previous research, the red flesh color trait in watermelon was found to be controlled by a single recessive gene (Henderson et al., 1998, Bang et al., 2010). To analyze the genetic inheritance patterns of red flesh color in watermelon, two F_2 populations and one BC population were tested in our research. The segregation of these three populations further verified that one major recessive gene controlled red flesh color and lycopene content. The segregation of flesh color in Pop. 1 was 9:3:4 in both 2013 (Liu et al., 2015) and 2014, and we propose that the gene for pale yellow (*Py*) flesh in watermelon is epistatic to the canary yellow–related gene based on the appearance of pale yellow flesh color in Pop. 1. The homozygous recessive allele of the *py* gene could inhibit the formation of canary yellow, yielding

a pale yellow flesh color. The results from the HPLC data of Pop. 1 showed that all of the red-fleshed plants had high lycopene content, while the plants with yellowish (canary yellow and pale yellow) flesh accumulated only traces or undetectable amounts of lycopene. The skewed distribution of lycopene content in Pop. 1 implied that a major effect gene controlled this trait.

Both LSW-177 and garden parent have red flesh color in their mature fruit, but their segregation was quite different when they were crossed with the white-fleshed PI 186490. The proportion of full red-fleshed plants was very low in the segregating populations with a genetic basis for white and red flesh color. Only two full red-fleshed plants were detected in Pop. 2, and in Pop. 3, there were approximately 22 full red fleshed individuals out of 116 plants. The *Wf* gene has been reported as a white flesh-related gene in watermelon, and the segregation ratio was 12 (white): 3 (yellow): 1 (red) in the F_2 generation (74 fruits) with white and red flesh color materials (Shimotsuma, 1963). In the research of Zhang and colleagues, watermelon lines 97103 (red-fleshed) and PI 296341-FR (light yellowish green flesh color, which is regarded as a white-fleshed parental line) were used to form F_2 and F_9-RIL populations. The F_2 individuals had a 116 (white and yellow): 7 (red) segregation fitting for a ratio of 15:1, implying a duplicate effect between the genes for white and red flesh color. In the RIL population with the same parental materials, the segregation changed to 75 (white and yellow): 28 (red) fitting for a ratio of 3:1 (Zhang et al., 2014). Another *Wf* gene investigation was reported by Gusmini and Wehner (2006), but in this study, COS, with pale yellow flesh color, was used as white-fleshed material to cross with a canary yellow flesh line, NC-517. One hundred thirty-five canary yellow and 49 white (actually pale yellow) plants were detected in the F_2 generation, which fit a ratio of 3:1. In
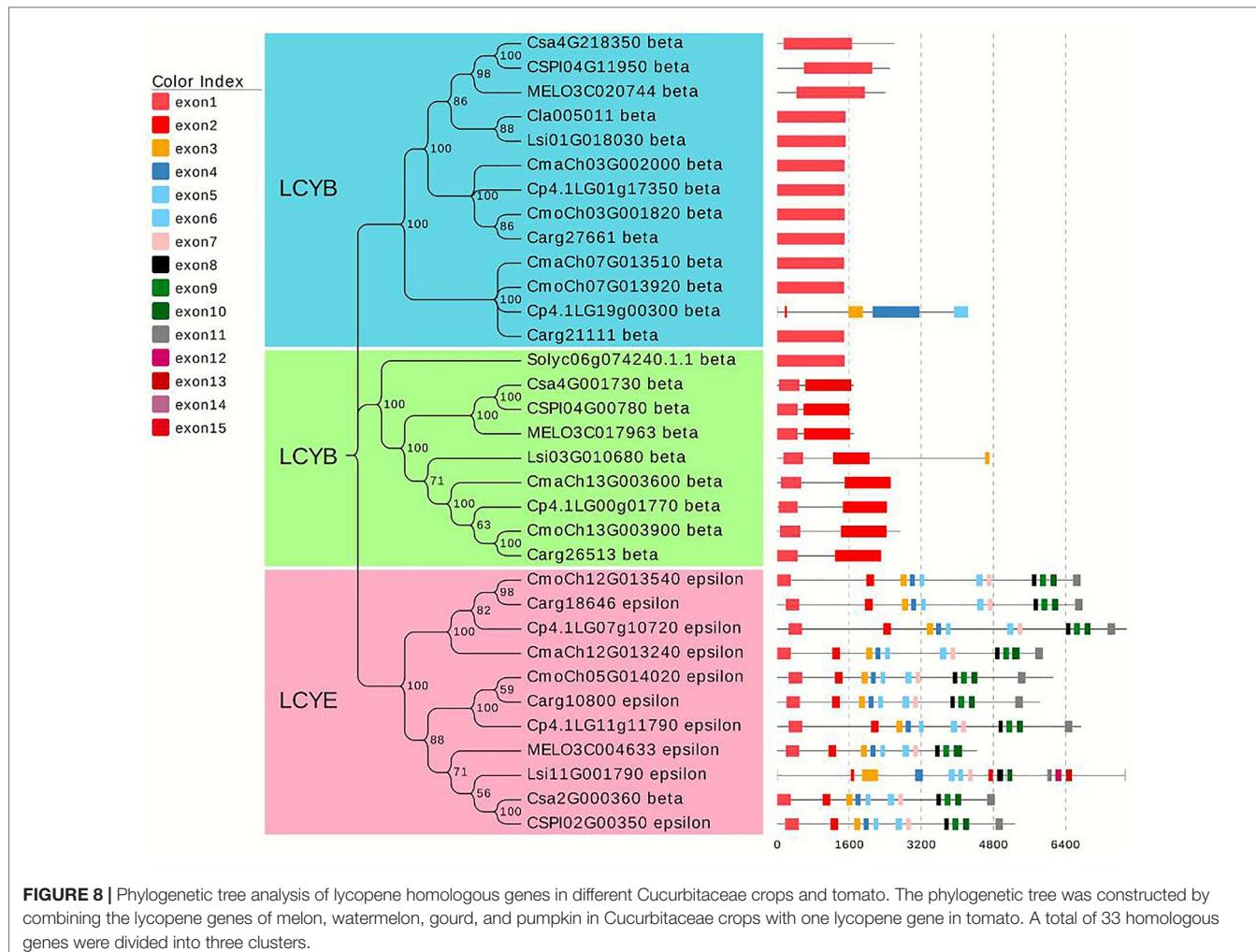
FIGURE 8 | Phylogenetic tree analysis of lycopene homologous genes in different Cucurbitaceae crops and tomato. The phylogenetic tree was constructed by combining the lycopene genes of melon, watermelon, gourd, and pumpkin in Cucurbitaceae crops with one lycopene gene in tomato. A total of 33 homologous genes were divided into three clusters.

our research, the segregation ratio was somewhat different from that of the previous study, which could be explained by the use of different parental lines. PI 186490, with a pure white flesh color, was different from COS and PI 296341-FR. According to the genotyping data with the gene marker *WII04EBsaHI-6* in Pop. 2 and Pop. 3, plants with a red or red-mixed white flesh color had the same alleles as those of the red parental materials and other red accessions. The ratio of red alleles and nonred alleles was 1:3 and 1:1 for Pop. 2 and Pop. 3, respectively. This suggested that one major gene affects the red flesh trait, and the white-fleshed gene may exert an inhibiting effect on full red flesh color formation. The white flesh color trait was incomplete dominant to the red flesh trait based on the fact that the two F_1 generations of Pop. 2 and Pop. 3 were yellowish mixed with white, not full white. The white-fleshed trait would be more complicated than pale yellow and red, which still needs further investigation.

Orange-fleshed plants were segregated in Pop. 3, while none were detected in Pop. 2, suggesting that the pigment compositions in LSW-177 and garden parent were quite different from each other; some carotenoid-containing orange flesh or a mixture of red and yellowish pigments accumulated in the mature flesh of garden parent.

The Locus Related to Red Flesh Color and the Lycopene Content Trait

The carotenoid biosynthesis pathway in plants has been thoroughly investigated (Cazzonelli and Pogson, 2010), but little research was focused on QTL analysis of lycopene accumulation in watermelon. Red tomato is also a high-lycopene-content plant, which reportedly undergoes pigment development similar to that of watermelon (Grassi et al., 2013); however, the two fruits are still suggested to have different regulatory systems in their carotenoid biosynthesis pathways. The genes in the carotenoid biosynthesis pathway perform through different action modes to regulate carotenoid accumulation and flesh color formation. Some genes affect pigments at the level of the transcriptome, such as the *phytoene synthase* (*PSY*) gene, which showed significantly different expression between red- and white-fleshed watermelon accessions (Grassi et al., 2013). On the

other hand, some genes did not show significant differences in expression among fruits with different flesh colors but still exhibited carotenoid accumulation variance based on enzyme activity; for example, the *CmOr* and *BoOr* genes acted as the main functional genes for β-carotene accumulation in melon and cauliflower (Lu et al., 2006; Tzuri et al., 2015).

By performing stepwise increases in mapping population sizes, marker numbers, and multiple genetic populations, a major effective QTL and candidate gene related to lycopene content and red flesh color was refined in a narrow region of chromosome 4. The two trait-related QTLs shared the same region, and the red flesh color gave a high correlation with the lycopene content in the F_2 generation for Pop. 1. Based on the preliminary mapping information, new CAPS markers were developed to narrow down the region from 392,077 bp to 41,233 bp. To verify the stability of the QTL we identified, Pop. 2 and Pop. 3 were also used in our research. The same QTL region and markers showed a high detection efficacy for red flesh color or lycopene content in the two populations through linkage analysis and MAS.

In this study, *Cla005011* seemed to be the best candidate gene for lycopene accumulation and red flesh color formation. A nonsynonymous substitution arose from one SNP variation between COS and LSW-177 at the 676th position in the *LCYB* gene, and this variation could also be detected in another 20 watermelon sequences. We also performed a transcriptome analysis of COS and LSW-177 at different time points throughout the whole growth period with the RNA of pulp, the expression level of *Cla005011* was not the main reason for red and nonred flesh color formation. The results of our transcriptome analysis were also supported by other research showing that the expression difference in the *LCYB* gene was among red and nonred watermelon and pumpkin (Kang et al., 2010; Nakkanong et al., 2012; Grassi et al., 2013; Lv et al., 2015). This might indicate that the accumulation of lycopene was not dependent on the expression of *LCYB*. It was also likely that the difference in protein levels or functionality may regulate the color of the fruit (Wang et al., 2016). This seemed contradictory when we combined the results of QTL mapping and transcriptome analysis. In liverwort, the results of functional identification also proved that the *LCYB* gene had a lycopene degradation capability to produce β-carotene at the enzyme activity level (Takemura et al., 2014). It is reasonable to speculate that *LCYB* may regulate lycopene metabolism through protein level. The nonsynonymous SNP locus in *Cla005011* in the 676th coding region (**Figure 5** and **Figure S3**) may be the key site causing the change in enzyme activity, which may still need further investigation. Although the red flesh color or lycopene content trait was recessive to the pale yellow and white flesh colors, some mixed color individuals still had lycopene accumulation in their mature fruits. This suggested the existence of some other loci acting as regulatory factors or functional genes in addition to the *LCYB* gene, which affected the lycopene accumulation in watermelon.

MAS Technology for Flesh Color in Watermelon

Flanking markers with small mapping intervals have the potential to increase MAS efficacy by reducing errors during selection. Two kinds of assisted selection markers for red flesh color in watermelon were generated by Bang et al. (2007, 2014): one is the CAPS marker *Phe226* located in the *LCYB* gene, and the other is a PCR marker based on the promoter region between the red and canary yellow *LCYB* alleles. According to our results, the CAPS marker *WII04EBsaHI-6* and *Phe226* shared the same location, and we also implemented a forward genetics strategy to demonstrate the significance of *LCYB* in lycopene accumulation in watermelon (red to nonred). Restriction enzyme digestion of *W04EII08-38*, *WII04EKpnI-1*, and *WII04E07-40* (*Mlu*I, *Kpn*I, and *Mbo*II) was more cost-effective than that of *WII04EBsaHI-6* and *Phe226* (*Bsa*HI). For other flesh colors, digestion was still weak according to a previous study. In our research, we developed two CAPS markers that cosegregated with yellowish flesh color both in the individuals of the genetic population and those in the natural panel. Two yellowish, three white, and one pink-fleshed watermelon accessions did not match with the MAS results, and the sequence comparison also showed that PI 248178 had a white flesh color but was encoded by the same *LCYB* sequence as that of the red accessions. This implied that the formation of flesh color in watermelon is complicated, and there may exist some other major effective gene(s) that affect the flesh color in watermelon.

The amount of pigment accumulation led to varying degrees of watermelon flesh color exhibited in one color system, such as pink to red and pale yellow to canary yellow. In our research, different shades of flesh color in one color system validate the same MAS results. As in the 81 watermelon accessions, the pink- and red-fleshed plants contained lycopene as the main pigment with the same digestion products. Based on these results, we speculated that the main effective genes could determine the formation of the red color system (pink gradually changing to red) and the yellow color system (from pale yellow to canary yellow). However, for each color system, other genes may regulate the amount of pigment accumulation to direct the formation of different shades of flesh color.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

CW, LS, and FL conceived and designed the experiments. AQ, XF, SL, and PG performed the field experiments. WC performed the data analysis and wrote the manuscript. AD offered the germplasm. All authors read and approved the final manuscript. It is worth noting that CW and AQ are co–first authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01240/full#supplementary-material

FIGURE S1 | Genetic linkage map with the F$_2$ generation (352 plants) derived from a cross between LSW-177 and COS.
The shaded area on linkage group 4 indicates the location of QTLs associated with red flesh color traits (*LCYB4.1*).
Chr: Chromosome

FIGURE S2 | Genetic linkage map with the F$_2$ generation (359 plants) derived from a cross between garden female and PI 186490.
Chr: Chromosome
The shaded area on linkage group 4 indicates the location of QTLs associated with red flesh color trait.

FIGURE S3 | The sequence alignment of *LCYB* gene with 24 watermelon accessions.

FIGURE S4 | Marker analysis for red, yellowish and white flesh color watermelon accessions using CAPS marker *WII04EBsaHI-6* and *WII04E08-38*.
Supplementary Figure S4a was the genotyping results of CAPS markers *WII04EBSAHI-6*; figure S4b was the genotyping results of CAPS markers *WII04E08-38*.
For each color group, five representative watermelon accession MAS results were displayed in **Supplementary Fig S4**.
Lane 1, 7, 13, 19 and 25 is the D2000 plus DNA marker. From the top to the bottom was 5,000, 3,000, 2,000, 1,000, 750, 500, 250, 100 bp fragments, respectively.
From lane 2 to 6 are five red flesh color watermelon accessions; from lane 8 to 12 are five yellowish flesh color watermelon accessions; from lane 14 to 18 are five pink flesh color watermelon accessions; from lane 20 to 24 are five white flesh color watermelon accessions.

FIGURE S5 | Marker analysis for red, yellowish and white flesh color watermelon accessions using CAPS marker *WII04EKpnI-1* and *WII04E07-40*.
Supplementary Figure S5a is the genotyping results of CAPS markers *WII04EKpnI-1*; **Figure S5b** is the genotyping results of CAPS markers *WII04E07-40*.
For each color group, five representative watermelon accessions MAS results are displayed in **Figure S5**. Lane 1, 7, 13, 19 and 25 are the D2000 plus DNA marker. From the top to the bottom are 5,000, 3,000, 2,000, 1,000, 750, 500, 250, 100 bp fragments, respectively. From lane 2 to 6 are five red flesh color watermelon accessions; from lane 8 to 12 are five yellowish flesh color watermelon accessions; from lane 14 to 18 are five pink flesh color watermelon accessions; from lane 20 to 24 are five white flesh color watermelon accessions.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1900). Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410. doi: 10.1016/S0022-2836(05)80360-2

Bang, H., Kim, S., Leskovar, D., and King, S. (2007). Development of a co-dominant 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

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(4), 362–368. doi: 10.21273/JASHS.135.4.362

Bang, H., Yi, G., Kim, S., Leskovar, D., and Patil, B. S. (2014). Watermelon lycopene β-cyclase: promoter characterization leads to the development of a PCR marker for allelic selection. *Euphytica* 200 (3), 363–378. doi: 10.1007/s10681-014-1158-5

Bo, K. L., Song, H., Shen, J., Qian, C. T., Staub, J. E., Simon, P. W., et al. (2011). Inheritance and mapping of the ore gene controlling the quantity of β-carotene in cucumber (*Cucumis sativus* L.) endocarp. *Mol. Breed.* 30, 335–344. doi: 10.1007/s11032-011-9624-4

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 (12), 146. doi: 10.1007/s11032-017-0747-0

Campbell, R., Pont, S. D. A., Morris, J. A., McKenzie, G., Sharma, S. K., Hedley, P. E., et al. (2014). Genome-wide QTL and bulked transcriptomic analysis reveals new candidate genes for the control of tuber carotenoid content in potato (*Solanum tuberosum* L.). *Theor. Appl. Genet.* 127 (9), 1917–1933. doi: 10.1007/s00122-014-2349-0

Capel, C., del Carmen, A. F., Alba, J. M., Lima-Silva, V., Hernandez-Gras, F., Salinas, M., et al. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. *Theor. Appl. Genet.* 128 (10), 2019–2035. doi: 10.1007/s00122-015-2563-4

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25 (15), 1972–1973. doi: 10.1093/bioinformatics/btp348

Cazzonelli, C. I., and Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15 (5), 266–274. doi: 10.1016/j.tplants.2010.02.003

Chayut, N., Yuan, H., Ohali, S., Meir, A., Yeselson, Y., Portnoy, V., et al. (2015). A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with orange allelic variation and fruit beta-carotene accumulation in melon fruit. *BMC Plant Biol.* 15, 247. doi: 10.1186/s12870-015-0661-8

Cuevas, H. E., Staub, J. E., Simon, P. W., Zalapa, J. E., and McCreight, J. D. (2008). Mapping of genetic loci that regulate quantity of beta-carotene in fruit of US Western Shipping melon (*Cucumis melo* L.). *Theor. Appl. Genet.* 117, 1345–1359. doi: 10.1007/s00122-008-0868-2

Danin-Poleg, Y., Reis, N., and Tzuri, G. (2001). Development and characterization of microsatellite in *Cucumis*. *Theor. Appl. Genet.* 102, 61–72. doi: 10.1007/s001220051618

Fazio, G., Staub, J. E., and Chung, S. M. (2002). Development and characterization of PCR markers in cucumber. *HortScience* 127, 545–557. doi: 10.21273/JASHS.127.4.545

Feng, D., Ling, W. H., and Duan, R. D. (2010). Lycopene suppresses LPS-induced NO and IL-6 production by inhibiting the activation of ERK, p38MAPK, and NF-k B in macrophages. *Inflamm. Res.* 59, 115–121. doi: 10.1007/s00011-009-0077-8

Fernandez-Silva, I., Eduardo, I., Blanca, J., Esteras, C., Pico, B., Nuez, F., et al. (2008). Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). *Theor. Appl. Genet.* 118, 139–150. doi: 10.1007/s00122-008-0883-3

Galpaz, N., Gonda, I., Shem-Tov, D., Barad, O., and Katzir, N. (2018). Deciphering genetic factors that determine melon fruit-quality traits using rna-seq–based high-resolution QTL and EQTL mapping. *Plant J.* 94 (1), 169–191. doi: 10.1111/tpj.13838

Gao, L., Zhao, W. H., Qu, H. O., Wang, Q. S., and Zhao, L. X. (2016). The yellow-fruited tomato 1 (*yft1*) mutant has altered fruit carotenoid accumulation and reduced ethylene production as a result of a genetic lesion in *ethylene insensitive2*. *Theor. Appl. Genet.* 129, 717–728. doi: 10.1007/s00122-015-2660-4

Gonzalo, M. J., Oliver, M., Garcia-Mas, J., Monfort, A., Dolcet-Sanjuan, R., Katzir, N., et al. (2005). Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). *Theor. Appl. Genet.* 110, 802–811. doi: 10.1007/s00122-004-1814-6

Grassi, S., Piro, G., Lee, J. M., Zheng, Y., Fei, Z. J., Dalessandro, G., et al. (2013). Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. *BMC Genomics* 14 (1), 781. doi: 10.1186/1471-2164-14-781

Guo, S. G., Zhang, J. G., Sun, H. H., Salse, J., Lucas, W. J., Zhang, H. Y., et al. (2012). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45 (1), 51–U82. doi: 10.1038/ng.2470

Gusmini, G., and Wehner, T. C. (2006). Qualitative inheritance of rind pattern and flesh color in watermelon. *J. Hered.* 97, 177–185. doi: 10.1093/jhered/esj023

Han, F. Q., Yang, C., Fang, Z. Y., Yang, L. M., Zhuang, M., Lv, H. H., et al. (2015). Inheritance and InDel markers closely linked to petal color gene (*cpc-1*) in *Brassica oleracea*. *Mol. Breed.* 35 (8), 160. doi: 10.1007/s11032-015-0354-x

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

Harel-Beja, R., Tzuri, G., Portnoy, V., Lotan-Pompan, M., Lev, S., Cohen, S., et al. (2010). A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. *Theor. Appl. Genet.* 121, 511–533. doi: 10.1007/s00122-010-1327-4

Henderson, W., Scott, G., and Wehner, T. C. (1998). Interaction of flesh color genes in watermelon. *J. Hered.* 89, 50–53. doi: 10.1093/jhered/89.1.50

Ivanov, B. N. (2014). Role of ascorbic acid in photosynthesis. *Biochem. (Mosc.)* 79, 282–289. doi: 10.1134/S0006297914030146

Joobeur, T., Gusmini, G., Zhang, X., Levi, A., Xu, Y., Wehner, T. C., et al. (2006). Construction of a watermelon BAC library and identification of SSRs anchored to melon or *Arabidopsis* genomes. *Theor. Appl. Genet.* 2, 1553–1562. doi: 10.1007/s00122-006-0258-6

Kang, B. S., Zhao, W. E., Hou, Y. B., and Tian, P. (2010). Expression of carotenogenic genes during the development and ripening of watermelon fruit. *Sci. Hortic. (Amsterdam)* 124, 368–375. doi: 10.1016/j.scienta.2010.01.027

Kapustin, Y., Souvorov, A., Tatusova, T., and Lipman, D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol. Direct* 3 (1), 20. doi: 10.1186/1745-6150-3-20

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi: 10.1093/molbev/msy096

Latowski, D., Kuczynska, P., and Strzalka, K. (2011). Xanthophyll cycle—a mechanism protecting plants against oxidative stress. *Redox Rep.* 16, 78–90. doi: 10.1179/174329211X13020951739938

Liu, C. H., Zhang, H. Y., Dai, Z. Y., Liu, X., Liu, Y., Deng, X. X., et al. (2012). Volatile chemical and carotenoid profiles in watermelons [*Citrullus vulgaris* (Thunb.) Schrad (Cucurbitaceae)] with different flesh colors. *Food Sci. Biotechnol.* 21 (2), 531–541. doi: 10.1007/s10068-012-0068-3

Liu, H., Jiao, J. Q., Liang, X. J., Liu, J., Meng, H. W., Chen, S. X., et al. (2016). Map-based cloning, identification and characterization of the *w* gene controlling white immature fruit color in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 129, 1247–1256. doi: 10.1007/s00122-016-2700-8

Liu, S., Gao, P., Wang, X. Z., Davis, A. R., Baloch, A. M., and Luan, F. S. (2015). Mapping of quantitative trait loci for lycopene content and fruit traits in *Citrullus lanatus*. *Euphytica* 202 (3), 411–426. doi: 10.1007/s10681-014-1308-9

Lu, H. W., Miao, H., Tian, G. L., Wehner, T. C., Gu, X. F., and Zhang, S. P. (2015). Molecular mapping and candidate gene analysis for yellow fruit flesh in cucumber. *Mol. Breed.* 35, 64. doi: 10.1007/s11032-015-0263-z

Lu, S., Van Eck, J., Zhou, X., Lopez, A. B., O'Halloran, D. M., Cosman, K. M., et al. (2006). The cauliflower *Or* gene encodes aDnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. *Plant Cell* 18, 3594–3605. doi: 10.1105/tpc.106.046417

Luan, F. S., Delannay, I., and Staub, J. E. (2008). Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. *Euphytica* 164, 445–461. doi: 10.1007/s10681-008-9699-0

Lv, P., Li, N., Liu, H., Gu, H. 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

Mortensen, A., Skibsted, L. H., Sampson, J., Rice-Evans, C., and Everett, S. A. (1997). Comparative mechanisms and rates of free radical scavenging by carotenoid antioxidants. *FEBS Lett.* 418, 91–97. doi: 10.1016/S0014-5793(97)01355-0

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 (3), 269–283. doi: 10.1016/j.cj.2015.01.001

Nakkanong, K., Yang, J. H., and Zhang, M. F. (2012). Carotenoid accumulation and carotenogenic gene expression during fruit development in novel interspecific inbred squash lines and their parents. *Agric. Food Chem.* 60 (23), 5936–5944. doi: 10.1021/jf3007135

Perkins-Veazie, P., Collins, J. K., Davis, A. R., and Roberts, W. (2006). Carotenoid content of 50 watermelon cultivars. *Agric. Food Chem.* 54, 2593–2597. doi: 10.1021/jf052066p

Shimotsuma, M. (1963). Cytogenetical studies in the genus *Citrullus*. VI. Inheritance of several characters in watermelons. *Japan. J. Breed.* 13, 235–240. doi: 10.1270/jsbbs1951.13.235

Silberstein, L., Kovalski, I., Brotman, Y., Perin, C., Dogimont, C., Pitrat, M., et al. (2003). Linkage map of *Cucumis melo* including phenotypic traits and sequence-characterized genes. *Genome* 46, 761–773. doi: 10.1139/g03-060

Sun, L., Yuan, B., Zhang, M., Wang, L., Cui, M. M., Wang, Q., et al. (2012). Fruit-specific RNAi-mediated suppression of *SlNCED1* increases both lycopene and β-carotene contents in tomato fruit. *J. Exp. Bot.* 63 (8), 3097–3108. doi: 10.1093/jxb/ers026

Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., and Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor. Appl. Genet.* 128 (5), 851–864. doi: 10.1007/s00122-015-2475-3

Tadmor, Y., King, S., Levi, A., Davis, A. R., Meir, A., Wasserman, B., et al. (2005). Comparative fruit coloration in watermelon and tomato. *Food Res. Int.* 38, 837–841. doi: 10.1016/j.foodres.2004.07.011

Takemura, M., Maoka, T., and Misawa, N. (2014). Carotenoid analysis of a liverwort *Marchantia polymorpha* and functional identification of its lycopene β- and ε-cyclase genes. *Plant Cell Physiol.* 55, 194–200. doi: 10.1093/pcp/pct170

Tzuri, G., Zhou, X. J., Chayut, N., Yuan, H., Portnoy, V., Meir, A., et al. (2015). A 'golden' SNP in CmOr governs the fruit flesh color of melon (*Cucumis melo*). *Plant J.* 82 (2), 267–279. doi: 10.1111/tpj.12814

Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93 (1), 77–78. doi: 10.1093/jhered/93.1.77

Wang, L. Y., Meng, X., Yang, D. Y., Ma, N. N., Wang, G. D., and Meng, Q. W. (2014). Overexpression of tomato GDP-L-galactose phosphorylase gene in tobacco improves tolerance to chilling stress. *Plant Cell Rep.* 33 (9), 1441–1451. doi: 10.1007/s00299-014-1627-2

Wang, N., Liu, S., Gao, P., and Luan, F. S. (2016). Developmental changes in gene expression drive accumulation of β-lycopene and carotene in watermelon. *J. Am. Soc. Hortic. Sci.* 141 (5), 434–443. doi: 10.21273/JASHS03690-16

Yi, K., Xu, X. L., Lu, X. Y., Xu, Y., Xiao, L. T., Wang, Y. J., et al. (2003). Construction of molecular genetic map of watermelon by SSR and ISSR technology. *J. Hunan Agric. Univ. (Nat. Sci.)* 29, 333–347. doi: 10.13331/j.cnki.jhau.2003.04.016

Young, A. J. (1991). The photoprotective role of carotenoids in higher plants. *Physiol. Plant* 83 (4), 702–708. doi: 10.1111/j.1399-3054.1991.tb02490.x

Yuan, L. P., Lu, X. Q., Liu, W. G., Zhao, J. S., Lu, J. S., Yan, Z. H., et al. (2012). Measurement of lycopene contents in watermelon. *China Cucurbits Veg.* 25, 1–4. doi: 10.16861/j.cnki.zggc.2012.03.001

Zalapa, J. E., Staub, J. E., McCreight, J. D., Chung, S. M., and Cuevas, H. (2007). Detection of QTL for yield-related traits using recombinant inbred lines derived from exotic and elite US Western Shipping melon germplasm. *Theor. Appl. Genet.* 114, 1185–1201. doi: 10.1007/s00122-007-0510-8

Zhao, W. E., Kang, B. S., and Hu, G. Q. (2008). Advances in research on carotenoids in watermelon flesh. *J. Fruit Sci.* 25 (6), 908–915. doi: 10.13925/j.cnki.gsxb.2008.06.018

Zhang, J., Gong, G. Y., Guo, S. G., Ren, Y., Zhang, H. Y., and Xu, Y. (2014). Fine mapping of the flesh color controlling genes in watermelon (*Citrullus lanatus*). *Cucurbitaceae 2014 Proceedings*, 111–116.

Zhang, J., Guo, S., Ren, Y., Zhang, H., Gong, G., Zhou, M., et al. (2016). High-level expression of a novel chromoplast phosphate transporter *ClPHT4;2* is required

for flesh color development in watermelon. *New Phytol.* 213 (3), 1208–1221. doi: 10.1111/nph.14257

Zhang, H., Wang, H., Guo, S., Ren, Y., Gong, G., Weng, Y., et al. (2011). Identification and validation of a core set of microsatellite markers for genetic diversity analysis in watermelon, *Citrullus lanatus* Thunb. Matsum. & Nakai. *Euphytica* 186, 329–342. doi: 10.1007/s10681-011-0574-z

Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2018). Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47 (D1), D1128–D1136 doi: 10.1093/nar/gky944

Cucumber CsTRY Negatively Regulates Anthocyanin Biosynthesis and Trichome Formation When Expressed in Tobacco

Leyu Zhang[1†], Jian Pan[1†], Gang Wang[1], Hui Du[1], Huanle He[1], Junsong Pan[1*] and Run Cai[1,2*]

[1] School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [2] State Key Laboratory of Vegetable Germplasm Innovation, Tianjin, China

*Correspondence:
Junsong Pan
jspan71@sjtu.edu.cn
Run Cai
cairun@sjtu.edu.cn

[†]These authors have contributed
equally to this work

The development of trichomes (spines) on cucumber fruits is an important agronomic trait. It has been reported that two MYB family members, *CsMYB6* (*Csa3G824850*) and *CsTRY* (*Csa5G139610*) act as negative regulators of trichome or fruit spine initiation. To further study the functions of these two genes, we overexpressed them in tobacco, and found that the flowers and seed coats of transformants overexpressing *CsTRY* displayed an unexpected defect in pigmentation that was not observed in plants overexpressing *CsMYB6*. Moreover, the expression of key genes in the flavonoid synthesis pathway was repressed in *CsTRY* overexpressing plants, which resulted in the decrease of several important flavonoid secondary metabolites. In addition, CsTRY could interact with the AN1 homologous gene CsAN1 (Csa7G044190) in cucumber, which further confirmed that CsTRY not only regulates the development of fruit spines, but also functions in the synthesis of flavonoids, acting as the repressor of anthocyanin synthesis.

Keywords: cucumber, trichome, anthocyanin, *CsTRY*, *CsMYB6*, tobacco

INTRODUCTION

Cucumber (*Cucumis sativus* L.) is a horticultural crop that is consumed worldwide (Huang et al., 2009; Yundaeng et al., 2015), and trichomes (spines) on the fruit are considered as an important commodity trait (Zhang et al., 2010; Yang et al., 2014; Li et al., 2015; Pan et al., 2015). The cucumber fruit, a pepo that develops from the ovary and receptacle, is covered with a thick cuticle, tubercules and trichomes (spines) (Roth, 1977; Wang et al., 2015). In the model plant *Arabidopsis thaliana*, trichome developments is initiated by a ternary complex (GL1-GL3/EGL1-TTG1) from epidermal cells, which leads to the expression of *GL2* and *TRY* (Oppenheimer et al., 1991; Galway et al., 1994; Walker, 1999; Payne et al., 2000; Szymanski et al., 2000; Larkin et al., 2003; Zhang et al., 2003). The TRY protein moves into neighboring cells, where it competes with *GL1* for binding to *GL3/EGL3* and prevents differentiation of the cells into trichomes (Schellmann et al., 2002; Esch et al., 2003; Zhang et al., 2003). Further, the transcriptional complex can also regulate anthocyanin biosynthesis genes to mediate anthocyanin biosynthesis, including NADPH-dependent dihydroflavonol reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), and UDP-Glc:flavonoid 3′- O-glucosyltransferase (UF3GT). In *Artemisia annua*, an MYB family member, *AaMIXTA1*, can promotes trichome development and regulates cuticle biosynthesis. These reports suggest that, the genes that regulate trichome developments usually function in secondary metabolite biosynthesis (Yan et al., 2018).

In cucumber, *csgl1/mict/tbh* mutants produced microtrichomes, which means the *GL1* gene may be related to the development of trichomes. *csgl3/tril* mutants have a hairless phenotype, which means the *GL3* gene is related to the initiation of trichomes (Chen et al., 2014; Li et al., 2015; Pan et al., 2015; Zhao et al., 2015; Wang et al., 2016), and *CsTTG* is involved in the formation of fruit warts (Chen et al., 2016). It has been reported that *CsMYB6* and *CsTRY* act as negative regulators of trichome initiation, and they can reduce cucumber trichome density (Yang et al., 2018). However, the specific regulatory mechanisms are still unclear.

In this study, we overexpressed *CsTRY* and *CsMYB6* in tobacco (*Nicotiana tabacum L.*) and found the flowers and seed coats of *CsTRY* overexpressing transformants displayed an unexpected defect in pigmentation that was not found in *CsMYB6* overexpressing plants. Furthermore, the expression of key genes in the flavonoid synthesis pathway were repressed in *CsTRY* overexpressing plants. In addition, we determined the compound content in the anthocyanin synthesis pathway by LC-MS and found that the content of peonidin and several important flavonoid secondary metabolites was significantly decreased, which is consistent with the gene expression change. *CsTRY* could interact with the *AN1* homologous gene in cucumber. These results suggested that *CsTRY* not only regulates the development of fruit spines, but also functions in synthesis of flavonoids, acting as the repressor of anthocyanin synthesis.

MATERIALS AND METHODS

CsTRY and *CsMYB6* Construct and Plant Transformation

The full-length *CsTRY* and *CsMYB6* coding region was amplified and inserted in the SacI and PstI sites of the pCambia2300 vector, containing the CaMV 35S promoter. The *CsMYB6* and *CsTRY* overexpression constructs were used for tobacco transformation (Horsch et al., 1985). The primes used are TRY-F(ATGGACAATCATCGT), TRY-R(TCATCCTCTTCTTCT),MYB6-F(ATGGGAAGGTCTCCT), MYB6-R(TCAGAATCTCAGGAA).

Preparation of Nucleic Acids and cDNA Synthesis

Genomic DNA was isolated from *N. tabacum* young leaf material using a DNeasy plant mini kit (QIAGEN). Total RNAs were extracted from *N. tabacum* petal material of the wild-type and transgenic plants, respectively, using an RNA extraction kit (TRIzol Reagent, Invitrogen). First-strand cDNA was synthesized from 2 μg total RNA in a 20 μl reaction mixture with 0.5 μg oligo(dT)$_{15}$, 0.75 mM dNTPs, 10 mM dithiothreitol (DTT), and 100 U SuperScript II RNase H-reverse transcriptase (Invitrogen).

Semi-Quantitative RT-PCR and Real-Time RT-PCR Analysis

For expression analysis of different structural flavonoid genes, the petunia genes of interest, *PhCHS*, *PhCHI*, *PhF3H*, *PhF3'H*, *PhDFR*, and *PhANS* were blasted with *N. tabacum* EST database.

SYBR® Premix Ex Taq from TaKaRa was used for qPCR with an Applied Biosystems 7500 real-time PCR system (Applied Biosystems). The tobacco gene *EF1α* was used as an internal control (Zhang et al., 2009) in all qPCR reactions. Three biological replicates were performed for each experiment.

Photometric Determination of Anthocyanins

Petals of mature flowers were harvested, ground in liquid nitrogen to produce a fine powder, and then immediately freeze-dried, and stored at −80°C until use. Anthocyanins were detected as previously described (Zhang et al., 2009). All samples were measured as triplicates in three independent biological replicates. Error bars represent +SE.

Yeast Two-Hybrid Screen

We cloned the cDNA sequences of *CsTRY* (full-length) and fused it into the pGADT7 vectors. The ORF of *CsAN1*, a homolog gene of the *Arabidopsis PhANTHOCYANIN1* (*AN1*) (Spelt, 2000) was cloned and fused into the pGBKT7. All recombinant constructs were separately transformed into the yeast strain AH109. At least three independent experiments were performed, and the result of one representative experiment is shown.

Scanning Electron Microscopy

N. tabacum young leaf samples were fixed, washed, post fixed, dehydrated, coated (Chen et al., 2014), and observed using a Hitachi S-4700 scanning electron microscope with a 2-kV accelerating voltage.

Bimolecular Fluorescence Complementation (BiFC) Assay

To generate the BiFC constructs, the full-length cDNA sequences of *CsTRY* and *CsAN1* were cloned and fused with the pXY104 and pXY106 vectors (Yu et al., 2008; Liu and Howell, 2010). Tobacco (*N. tabacum*) leaves were used for co-expression studies as previously described (Schütze et al., 2009). The fluorescence signal was detected 2 to 4 days after infiltration, using an Olympus BX 51 fluorescence microscope to acquire fluorescent images. YFP (yellow fluorescent protein) imaging was performed at an excitation wavelength of 488 nm. CFP served as the internal control in all BiFC analyses. At least three independent replicates were performed, and the result of one representative experiment is shown.

Metabolite Profiling

The petals of transgene plants were grounded into a fine powder. Each 20 mg of fine powder was used for metabolite extraction prior to UHPLC-Q-TOF-MS analysis. The metabolite extraction was analyzed as previously described (Hu et al., 2018). The metabolites were annotated by searching the Personal Compound Database and Library (PCD/PCDL) (Hu et al., 2015), and by comparing the MS and MS/MS of the compounds in the Metlin database (Want, 2005) and the Massbank database (Horai et al., 2010). Data acquisition, metabolite annotation and

peak area extraction were performed with the Agilent softwares (Agilent Technologies Inc., Palo Alto, CA, USA), of MassHunter Acquisition 7.0, MassHunter Qualitative 7.0 and Mass Profinder 8.0, respectively. All measurements were performed in three replicates per genotype.

RESULTS

1. Flower Pigmentation and Trichome Distribution Were Affected in Transgenic Tobacco Plants Overexpressing *CsTRY*, but Not in Plants Overexpressing *CsMYB6*

To explore the function of *CsTRY* and *CsMYB6*, we used 35S promoter to regulate these two genes and overexpress them in tobacco by genetic transformation (**Figure 1A**). In transgenic 35S:CsTRY tobacco, there were clear phenotypic changes in petal pigmentation, which resulted in a complete loss of pigmentation and pure white petals. Moreover, there was a decrease in seed pigmentation, which was lighter than the wild type (**Figure 1B**).

Anthocyanin quantification results measured by spectrophotometer revealed that anthocyanin accumulation in the petals of 35S:CsTRY transgenic tobacco plants was clearly reduced, indicating that *CsTRY* may be negatively regulating the synthesis of tobacco anthocyanin (**Figure 1D**).

In addition, the morphology and quantity of glandular hairs of transgenic lines were observed by scanning election microscopy (**Figure 2**). It was found that the number of long stalked glandular hairs and the density of glandular hairs decreased, which is consistent with the known negative regulation of the epidermis. However, overexpression of *CsMYB6* in tobacco showed no difference in phenotypes, such as flower pigment and glandular density.

2. *CsTRY* Negatively Regulates the Synthesis of Anthocyanins by Suppressing the Expression of Genes in the Flavonoid Metabolic Pathway

To elucidate the molecular mechanisms involved in the marked decrease of anthocyanins in 35S:CsTRY transformants,
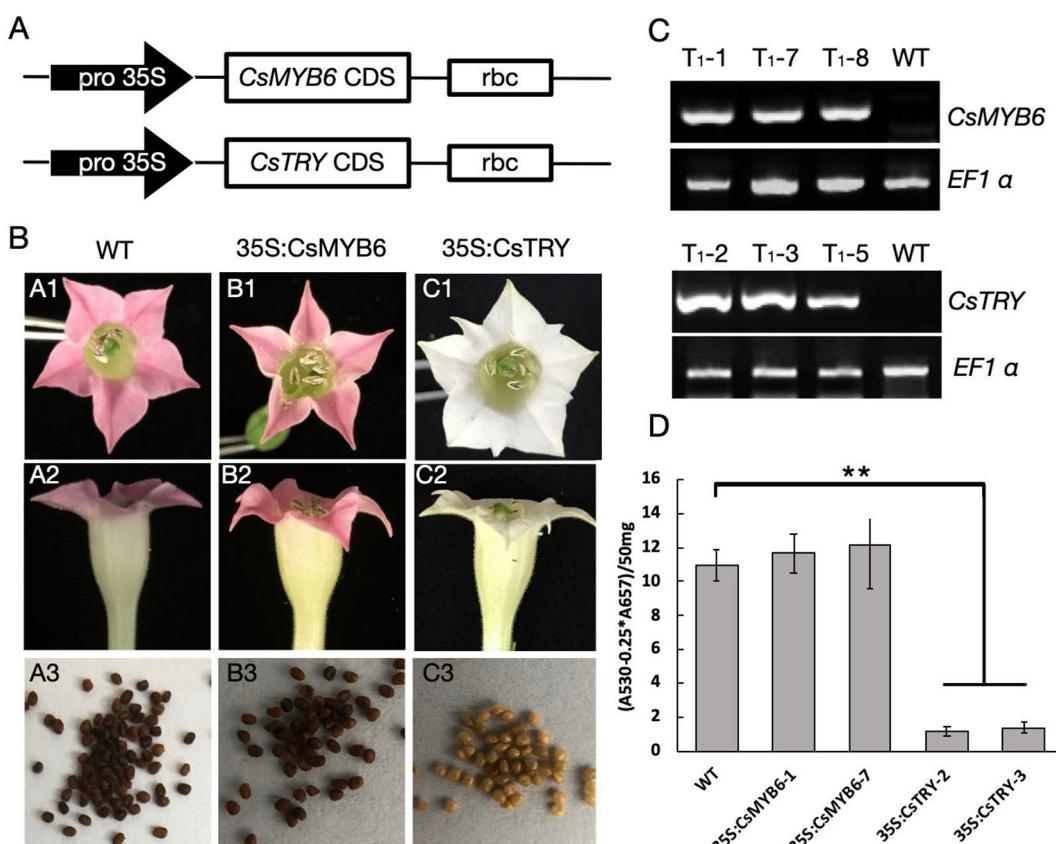
FIGURE 1 | The gene construct for ectopic expression of *CsTRY* and *CsMYB6* and changes in flower and seed pigmentation phenotype of transgenic tobacco. **(A)** Structure of the gene construct used to express. **(B)** *CsTRY* and *CsMYB6* expressing lines exhibiting different phenotype characteristics. A1–A3 is the petal and seeds of WT; B1–B3 is the petal and seeds of 35S:CsMYB6 transgenic tobacco plants; and C1–C3 is of 35S:CsTRY transgenic plants **(C)** Semi-quantitative RT-PCR analysis of *CsTRY* and *CsMYB6* expression levels in mature leaves of T1 generation plants. EF1α transcript abundance was used as a control. **(D)** Photometric determination of anthocyanin content in methanolic extracts of petals in tobacco lines *35S:CsTRY* (35S:CsTRY-2, 35S:CsTRY-3), *35S:CsMYB6* (35S:CsMYB6-1, 35S:CsMYB6-7) and the wild-type. A_{530}, absorption at 530 nm; A_{657}, absorption at 657 nm. Error bars represent +SE. Significant differences were determined according to Duncan's multiple range test (P < 0.05) or Student's t-test (**P < 0.01).
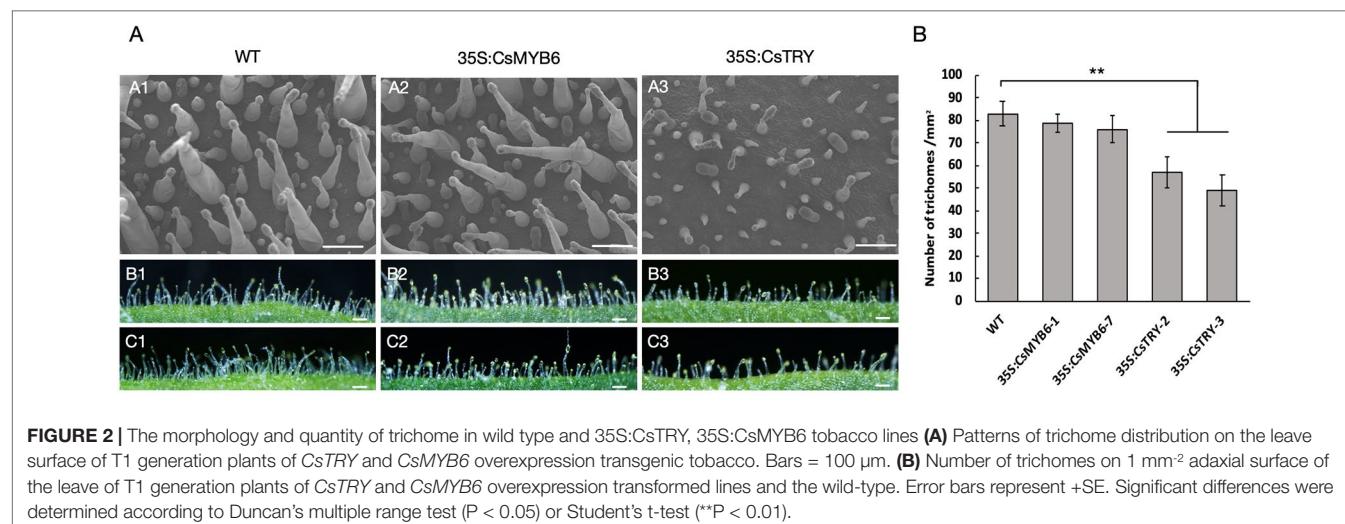
FIGURE 2 | The morphology and quantity of trichome in wild type and 35S:CsTRY, 35S:CsMYB6 tobacco lines **(A)** Patterns of trichome distribution on the leave surface of T1 generation plants of *CsTRY* and *CsMYB6* overexpression transgenic tobacco. Bars = 100 μm. **(B)** Number of trichomes on 1 mm^{-2} adaxial surface of the leave of T1 generation plants of *CsTRY* and *CsMYB6* overexpression transformed lines and the wild-type. Error bars represent +SE. Significant differences were determined according to Duncan's multiple range test (P < 0.05) or Student's t-test (**P < 0.01).

the transcript levels of seven key genes encoding the enzymes of the flavonoid pathway from the first stage to the third were measured in flowers by real-time RT-PCR analyses (**Figure 3**).

Based on the different responses of the seven genes to *CsTRY*, we divided them into three categories. The first type is Chalcone synthase (CHS) located upstream of the second stage of the flavonoid metabolic pathway, and its expression level is significantly inhibited; the second type contains Chalcone isomerase (CHI), flavanone 3-hydroxylas (F3H), and flavonoid 3′-hydroxylase (F3′H) located downstream of the second stage of the flavonoid metabolic pathway, and its expression level is upregulated to varying degrees. The third category is dihydroflavonol4-reductas (DFR), Anthocyanidin Synthase (ANS) and Anthocyanin 3-0-g1ucosyltransferase (3GT), which is directly related to the synthesis of anthocyanins and is clearly strongly inhibited.

Furthermore, we detected a variety of flavonoids in transgenic tobacco petals by LC-MS (**Figure 4**). We found that a number of secondary metabolites related to anthocyanins, such as (kaempferol-3-O-rhamnoside-7-O-rhamnoside and kaempferol-3-O-rutinoside-7-O-rhamnoside), naringenin hexoside, chalcone 2″-O-glucoside, anthocyanin (peonidin di-hexoside I, II, and III), were significantly reduced in transgenic plants.

This is consistent with the expression of genes. In the secondary stage of Anthocyanin synthesis, *CHS* catalyzes the stepwise condensation of three acetate units from malonyl-COA with p-coumaroyl-COA to yield tetrahydroxychalcone. *CHI* then catalyzes the stereospecific isornerization of the yellow-colored tetrahydroxychalcone to the colorless naringenin. Naringenin is converted to dihydrokaempferol by *F3H*. Thus, the decrease in the expression of *CHS* and the increase in the expression of *CHI, F3H* reduced the content of chalcone and trihydroxyflavanone. In the last stage of Anthocyanin synthesis, the dihydroflavonol is reduced to the flavan 9,4 cis-diol (leucoanthocyanidins) by dihydroflavone-4-reductase (DFR). Then Anthocyanin formed

by catalysis of *ANS* and *3GT*. Owing to the inhibited of the expression three genes, *DFR, ANS,* and *3GT*, the decrease of Anthocyanin is reasonable.

3. The CsTRY Protein Interacts With the Known Anthocyanin Synthesis Regulator *CsAN1*

The PhAN1 protein has previously been shown to be essential for anthocyanin synthesis in petunia. To further verify whether CsTRY can interact with known modulators of anthocyanin metabolism, we cloned the homologous gene of *PhAN1* in cucumber, *CsAN1*, through yeast two-hybrid and BiFC and found that CsTRY can interact with CsAN1 (**Figure 5**).

DISCUSSION

Trichomes are generally considered biofactories that produce secondary metabolites. The genes regulating unicellular trichome developments are usually related to anthocyanin synthesis (Jin and Martin, 1999). In *A. annua*, which has multicellular trichomes, genes involved in the development of trichomes also regulate the synthesis and transportation of secondary metabolites, such as artemisinin (Yan et al., 2018). In the cucumber trichome development mutants *csgl3/tril* and *csgl1/mict/tbh*, the genes that regulate anthocyanin synthesis are also differentially expressed, suggesting that although trichomes morphology and related genes differ between multicellular and unicellular trichomes, the mechanism of secondary metabolites coupled with epidermal hair development is conserved. Overexpression of *CsTRY* in tobacco can affect the flowering and seed coat color, but overexpression of *CsMYB6* does not. This result suggests that although both genes can regulate the density of trichomes in cucumber, the specific mechanism and range of the regulation may not be the same. In addition, *CsMYB6* was significantly downregulated in the cucumber hairless mutants *tril* and *mict*, but the expression of *CsTRY* did not change, which also indicated that the regulation
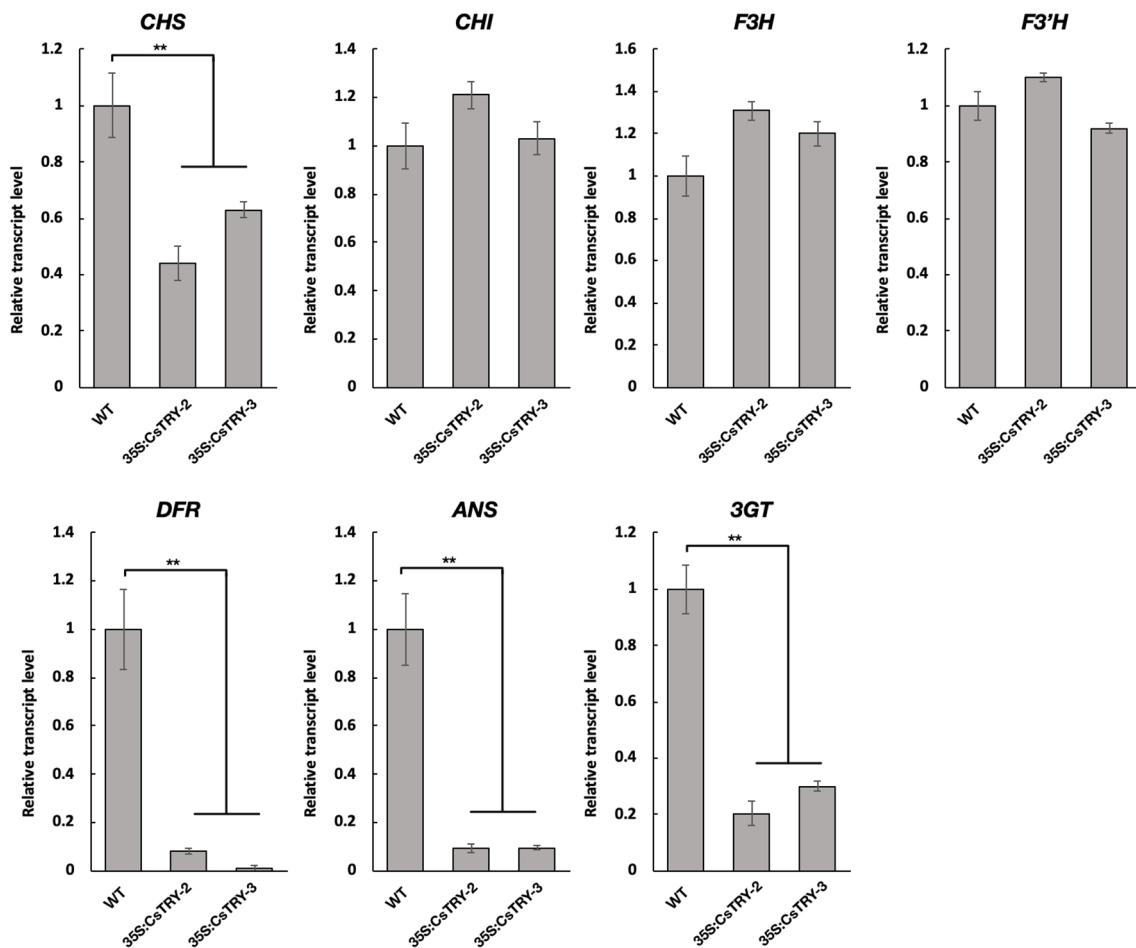
FIGURE 3 | Relative transcript levels of flavonoid structural genes in the petals of tobacco lines *35S:CsTRY* (35S:CsTRY-2, 35S:CsTRY-3) and wild type, with the *EF1α* gene as an internal control. Error bars represent +SE. Significant differences were determined according to Duncan's multiple range test (P < 0.05) or Student's t-test (**P < 0.01).
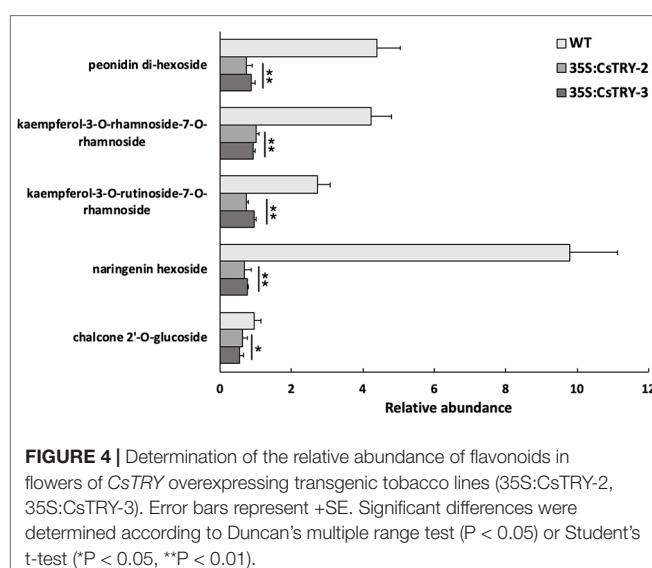
FIGURE 4 | Determination of the relative abundance of flavonoids in flowers of *CsTRY* overexpressing transgenic tobacco lines (35S:CsTRY-2, 35S:CsTRY-3). Error bars represent +SE. Significant differences were determined according to Duncan's multiple range test (P < 0.05) or Student's t-test (*P < 0.05, **P < 0.01).

patterns of *CsTRY* and *CsMYB6* may be different (Chen et al., 2014; Li et al., 2015; Zhao et al., 2015).

In previous study, overexpression *CsTRY* or *CsMYB6* in cucumber can decrease the density of fruit trichome and *CsTRY* is directly regulated by *CsMYB6*. However, overexpression of CsMYB6 inhibited rather than promoted, the expression of *CsTRY*, which means the relationship between *CsTRY* and *CsMYB6* is not simple (Yang et al., 2018). In this study, overexpression of *CsTRY* in tobacco can affect the number of trichome, but overexpression of *CsMYB6* does not. It can be inferred that *CsTRY*, a R3 MYB transcription factor in the relative downstream, was more directly related to glandular trichomes and metabolites, while *CsMYB6*, a R2R3 MYB transcription factor in the relative upstream, might be affected by other proteins in tobacco. Interestingly, overexpression *CsTRY* or *CsMYB6* in cucumber can decrease the density of fruit trichome rather than other organs (Yang et al., 2018), which means the regulation mechanism of fruit trichome is different from that of other organs. *SlMIXTA-like*, a R2R2MYB transcription factor

FIGURE 5 | Yeast two-hybrid analysis and BiFC **(A)** Activation activity of CsTRY to CsAN1 in yeast. **(B)** BiFC analysis of the physical interaction between CsAN1 (fused with the N-terminal fragment of YFP) and CsTRY (fused with the C-terminal fragment of YFP).
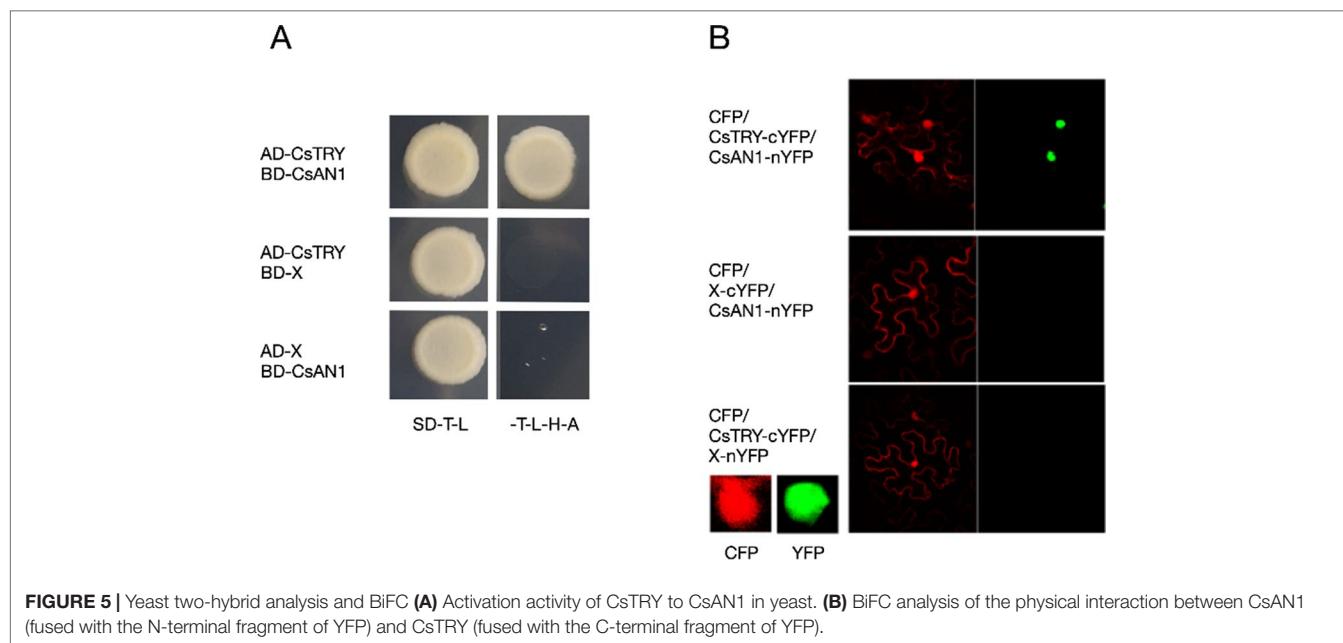
of tomato, regulates trichome formation on fruit surface, which also indicates that the formation of fruit trichome may be different from that of other organs (Ying et al., 2019). Therefore, overexpress *CsMYB6* in tobacco did not change the phenotype, suggesting that *CsMYB6* only plays a role in the regulation of cucumber fruit trichome, not other organs.

At present, there are three main ways to regulate anthocyanin synthesis: MYB-bHLH protein binary complex, MYB-WD40 protein binary complex, which is independent of bHLH transcription factor, and MYB-bHLH-WD40 protein ternary complex. The anthocyanin pathway in most plants is activated by MYB-bHLH-WD40 ternary complex (Broun, 2005; Antoine et al., 2010). In this study, we found that CsTRY can interact with the bHLH protein CsAN1. *AN1* is an important regulator involved in anthocyanin synthesis in *Petunia hybrida*. It is a homology of the structural gene DFR and can directly regulate the expression of *DFR* (Spelt, 2000). Without the research of WD40 protein, we speculate that *CsTRY* may function by forming a ternary complex of MYB-bHLH-WD40 protein or a binary complex of MYB-bHLH protein in tobacco. Moreover, *CsTRY* can interact with *CsAN1*, a homolog of the *Arabidopsis AN1* gene, indicating that it also acts as a negative regulator of anthocyanin synthesis in cucumber and has regulatory mechanisms similar to *Arabidopsis*.

In this study, the overexpression of *CsTRY* in tobacco greatly affected the synthesis of anthocyanin, indicating that *CsTRY* may have conserved function in cucumber. A number of transcription factors (*e.g., CsGl1/Mict/tbh, CsGl3/Tril, Tu* and *Ts*) play key roles in cucumber trichome (spine) differentiation and development. However, the functions of their homologues in *Arabidopsis* are irrelevant to trichome development (Yang et al., 2014; Zhao et al., 2015; Guo et al., 2018). Moreover, *CsTRY* can complement the *Arabidopsis try* mutant, whereas *CsMYB6* cannot complement the *gl1* mutant. These results suggest that the

regulation network of multicellular-trichome development may be partially consistent with unicellular-trichome development, but key genes, such as *CsGl3/Tril* and *CsGl1/Mict/tbh*, may have independent evolutionary pathways in the two different trichome types.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

HH conceptualized the research. JuP designed experiments. JiP participated in writing, editing, and revising the manuscript. RC conceptualized the research, designed and performed experiments. GW performed experiments and prepared figures. LZ performed experiments, analyzed the data, and wrote the manuscript. HD prepared the figures.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Antoine, B., Heim, M. A., Bertrand, D., Michel, C., Bernd, W., and LoC, L. (2010). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *Plant J.* 39 (3), 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Broun, P. (2005). Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. *Curr. Opin. Plant Biol.* 8 (3), 272–279. doi: 10.1016/j.pbi.2005.03.006

Chen, C., Liu, M., Jiang, L., Liu, X., Zhao, J., Yan, S., et al. (2014). Transcriptome profiling reveals roles of meristem regulators and polarity genes during fruit trichome development in cucumber (*Cucumis sativus* L.). *J. Exp. Bot.* 65 (17), 4943–4958. doi: 10.1093/jxb/eru258

Chen, C., Yin, S., Liu, X., Liu, B., Yang, S., Xue, S., et al. (2016). The WD-repeat protein CsTTG1 regulates fruit wart formation through interaction with the homeodomain-leucine zipper I protein Mict. *Plant Physiol.* 171 (2), 1156–1168. doi: 10.1104/pp.16.00112

Esch, J. J., Chen, M., Sanders, M., Hillestad, M., Ndkium, S., Idelkope, B., et al. (2003). A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in Arabidopsis. *Development* 130, 5885–5894. doi: 10.1242/dev.00812

Galway, M. E., Masucci, J. D., Lloyd, A. M., Walbot, V., Davis, R. W., and Schiefelbein, J. W. (1994). The TTG gene is required to specify epidermal cell fate and cell patterning in the Arabidopsis root. *Dev. Biol.* 166 (2), 0–754. doi: 10.1006/dbio.1994.1352

Guo, C., Yang, X., Wang, Y., Nie, J., Yang, Y., Sun, J., et al. (2018). Identification and mapping of ts (tender spines), a gene involved in soft spine development in *Cucumis sativus*. *Theor. Appl. Genet.* 131 (1), 1–12. doi: 10.1007/s00122-017-2954-9

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). Massbank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45 (7), 703–714. doi: 10.1002/jms.1777

Horsch, R., Hoffmann, J., Eichholtz, D., Rogers, S., and Fraley, R. (1985). A simple and general method for transferring genes into plants. *Science* 227 (4691), 1229–1231. doi: 10.1126/science.227.4691.1229

Hu, C., Tohge, T., Chan, S. A., Song, Y., Rao, J., Cui, B., et al. (2015). Identification of conserved and diverse metabolic shifts during rice grain development. *Sci. Rep.* 6, 20942. doi: 10.1038/srep20942

Hu, C., Zhao, H., Wang, W., Xu, M., Shi, J., Nie, X., et al. (2018). Identification of conserved and diverse metabolic shift of the stylar, intermediate and peduncular segments of cucumber fruit during development. *Int. J. Mol. Sci.* 19 (1). doi: 10.3390/ijms19010135

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

Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* 41 (5), 577–585. doi: 10.1023/A:1006319732410

Larkin, J. C., Brown, M. L., and Schiefelbein, J. (2003). How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis. *Annu. Rev. Plant Biol.* 54, 403–430. doi: 10.1146/annurev.arplant.54.031902.134823

Li, Q., Cao, C., Zhang, C., Zheng, S., Wang, Z., Wang, L., et al. (2015). The identification of *Cucumis sativus* Glabrous 1 (CsGL1) required for the formation of trichomes uncovers a novel function for the homeodomain-leucine zipper I gene. *J. Exp. Bot.* 66 (9), 2515–2526. doi: 10.1093/jxb/erv046

Liu, J. X., and Howell, S. H. (2010). bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. *Plant Cell* 22 (3), 782–796. doi: 10.1105/tpc.109.072173

Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. (1991). A MYB gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell* 67, 483–493. doi: 10.1016/0092-8674(91)90523-2

Pan, Y., Bo, K., Cheng, Z., and Weng, Y. (2015). The loss-of-function GLABROUS 3 mutation in cucumber is due to LTR-retrotransposon insertion in a class IV HD-ZIP transcription factor gene CsGL3 that is epistatic over CsGL1. *BMC Plant Biol.* 15, 302. doi: 10.1186/s12870-015-0693-0

Payne, C. T., Zhang, F., and Lloyd, A. M. (2000). Gl3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. *Genetics* 156 (3), 1349–1362. doi: 10.1016/S1474-6670(17)69558-2

Roth, I. (1977). Fruits of Cucurbitaceae. *Encyclopedia Plant Anat.* 10, 471–477. doi: 10.1016/S1474-6670(17)69558-2

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., et al. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. *EMBO (Eur. Mol. Biol. Organ.) J.* 21 (19), 5036–5046. doi: 10.1093/emboj/cdf524

Schütze, K., Harter, K., and Chaban, C. (2009). Bimolecular fluorescence complementation (BiFC) to study protein–protein interactions in living plant cells. *Methods Mol. Biol.* 479, 189–202. doi: 10.1007/978-1-59745-289-2_12

Spelt, C. (2000). Anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural Anthocyanin genes. *Plant Cell* 12 (9), 1619–1632. doi: 10.1105/tpc.12.9.1619

Szymanski, D. B., Lloyd, A. M., and Marks, M. D. (2000). Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. *Trends Plant Sci.* 5 (5), 214–219. doi: 10.1016/S1360-1385(00)01597-1

Walker, A. R. (1999). The transparent testa glabra1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell Online* 11 (7), 1337–1350. doi: 10.1105/tpc.11.7.1337

Wang, W., Zhang, Y., Xu, C., Ren, J., Liu, X., Black, K., et al. (2015). Cucumber ECERIFERUM1 (CsCER1), which influences the cuticle properties and drought tolerance of cucumber, plays a key role in VLC alkanes biosynthesis. *Plant Mol. Biol.* 87 (3), 219–233. doi: 10.1007/s11103-014-0271-0

Wang, Y. L., Nie, J. T., Chen, H. M., Guo, C. L., Pan, J., He, H. L., et al. (2016). Identification and mapping of Tril, a homeodomain-leucine zipper gene involved in multicellular trichome initiation in *Cucumis sativus*. *Theor. Appl. Genet.* 129 (2), 305–316. doi: 10.1007/s00122-015-2628-4

Want, E. J. (2005). Metlin: a metabolite mass spectral database. *Ther. Drug Monit.* 27 (6), 747. doi: 10.1097/01.ftd.0000179845.53213.39

Yan, T., Li, L., Xie, L., Chen, M., Shen, Q., Pan, Q., et al. (2018). A novel HD-ZIP IV/MIXTA complex promotes glandular trichome initiation and cuticle development in *Artemisia annua*. *New Phytol.* 218 (2), 567–578. doi: 10.1111/nph.15005

Yang, S., Cai, Y., Liu, X., Dong, M., Zhang, Y., Chen, S., et al. (2018). A CsMYB6-CsTRY module regulates fruit trichome initiation in cucumber. *J. Exp. Bot.* 69 (8), 1887–1902. doi: 10.1093/jxb/ery047

Yang, X., Zhang, W., He, H., Nie, J., Bie, B., Zhao, J., et al. (2014). Tuberculate fruit gene Tu encodes a C2 H2 zinc finger protein that is required for the warty fruit phenotype in cucumber (*Cucumis sativus* L.). *Plant J.* 78 (6), 1034–1046. doi: 10.1111/tpj.12531

Ying, S. Y., Su, M., Wu, Y., Zhou, L., et al. (2019). Trichome regulator *SlMIXTA-like* directly manipulates primary metabolism in tomato fruit. *Plant Biotechnol. J.*, 1–10. doi: 10.1111/pbi.13202

Yu, X., Li, L., Li, L., Guo, M., Chory, J., and Yin, Y. (2008). Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 105 (21), 7618–7623. doi: 10.1073/pnas.0802254105

Yundaeng, C., Somta, P., Tangphatsornruang, S., Chankaew, S., and Srinives, P. (2015). A single base substitution in BADH/AMADH is responsible for fragrance in cucumber (*Cucumis sativus* L.), and development of SNAP markers for the fragrance. *Theor. Appl. Genet.* 128 (9), 1881–1892. doi: 10.1007/s00122-015-2554-5

Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* 130 (20), 4859–4869. doi: 10.1242/dev.00681

Zhang, W., He, H., Guan, Y., Du, H., Yuan, L., Li, Z., et al. (2010). Identification and mapping of molecular markers linked to the tuberculate fruit gene in the

cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 120 (3), 645–654. doi: 10.1007/s00122-009-1182-3

Zhang, W., Ning, G., Lv, H., Liao, L., and Bao, M. (2009). Single MYB-type transcription factor AtCAPRICE: a new efficient tool to engineer the production of anthocyanin in tobacco. *Biochem. Biophys. Res. Commun.* 388 (4), 742–747. doi: 10.1016/j.bbrc.2009.08.092

Zhao, J. L., Pan, J. S., Guan, Y., Zhang, W. W., Bie, B. B., Wang, Y. L., et al. (2015). Micro-trichome as a class I homeodomain-leucine zipper gene regulates multicellular trichome development in *Cucumis sativus*. *J. Integr. Plant Biol.* 57, 925–935. doi: 10.1111/jipb.12345

Check for
updates

Gene Interactions Regulating Sex Determination in Cucurbits

Dandan Li[1], Yunyan Sheng[1], Huanhuan Niu[2] and Zheng Li[2]

[1] College of Horticulture and Landscape Architecture, Heilongjiang Bayi Agriculture University, Daqing, China,
[2] College of Horticulture, Northwest A&F University, Yangling, China

The family Cucurbitaceae includes many economically important crops, such as cucumber (*Cucumis sativus*), melon (*Cucumis melo*), watermelon (*Citrullus lanatus*), and zucchini (*Cucurbita pepo*), which share homologous gene pathways that control similar phenotypes. Sex determination is a research hotspot associated with yield and quality, and the genes involved are highly orthologous and conserved in cucurbits. In the field, six normal sex types have been categorized according to the distribution of female, male, or bisexual flowers in a given plant. To date, five orthologous genes involved in sex determination have been cloned, and their various combinations and expression patterns can explain all the identified sex types. In addition to genetic mechanisms, ethylene controls sex expression in this family. Two ethylene signaling components have been identified recently, which will help us to explore the ethylene signaling-mediated interactions among sex-related genes. This review discusses recent advances relating to the mechanism of sex determination in cucurbits and the prospects for research in this area.

Keywords: Cucurbitaceae, sex determination, ethylene, gene interaction, transcriptional regulation

INTRODUCTION OF SEX TYPES IN CUCURBITS

Flower development is the basis of fruit and seed production in plants. In angiosperms, ~90% of species have perfect flowers with separate stamens and carpels simultaneously. Compared with perfect or bisexual flowers, the differential development or selectivity arrest in the carpel or stamen in some species results in unisexual male or female flowers, respectively, leading to flower sex-type diversity (Tanurdzic and Banks, 2004). Various combinations or distributions of the three kinds of flowers produce hermaphroditic, dioecious, and monoecious plants, thus forming the sex-type diversity of plant. As in animals, the regulation in these floral developmental processes are defined as sex determination or sex differentiation. Sex determination in angiosperms has been well studied in recent years (reviewed by Tanurdzic and Banks, 2004; Lai et al., 2017; Pannell, 2017; Pawełkowicz et al., 2019a).

The family Cucurbitaceae comprises about 120 genera and 960 species, including many economically important crops such as cucumber (*Cucumis sativus*), melon (*Cucumis melo*), watermelon (*Citrullus lanatus*), zucchini (*Cucurbita pepo*), and pumpkins (*Cucurbita moschata*) (Bhowmick and Jha, 2015). The family Cucurbitaceae has abundant flower and plant sex types, and the regulation of sex determination can directly influence their yield and quality. Depending on the distribution or ratio of the three types of flowers produced in a plant (**Figure 1A**), the family Cucurbitaceae is classified into six phenotypes: monoecy, gynoecy, subgynoecy, androecy, andromonoecy, and hermaphrodite (**Figure 1B**). The most common sex type in *Cucumis*
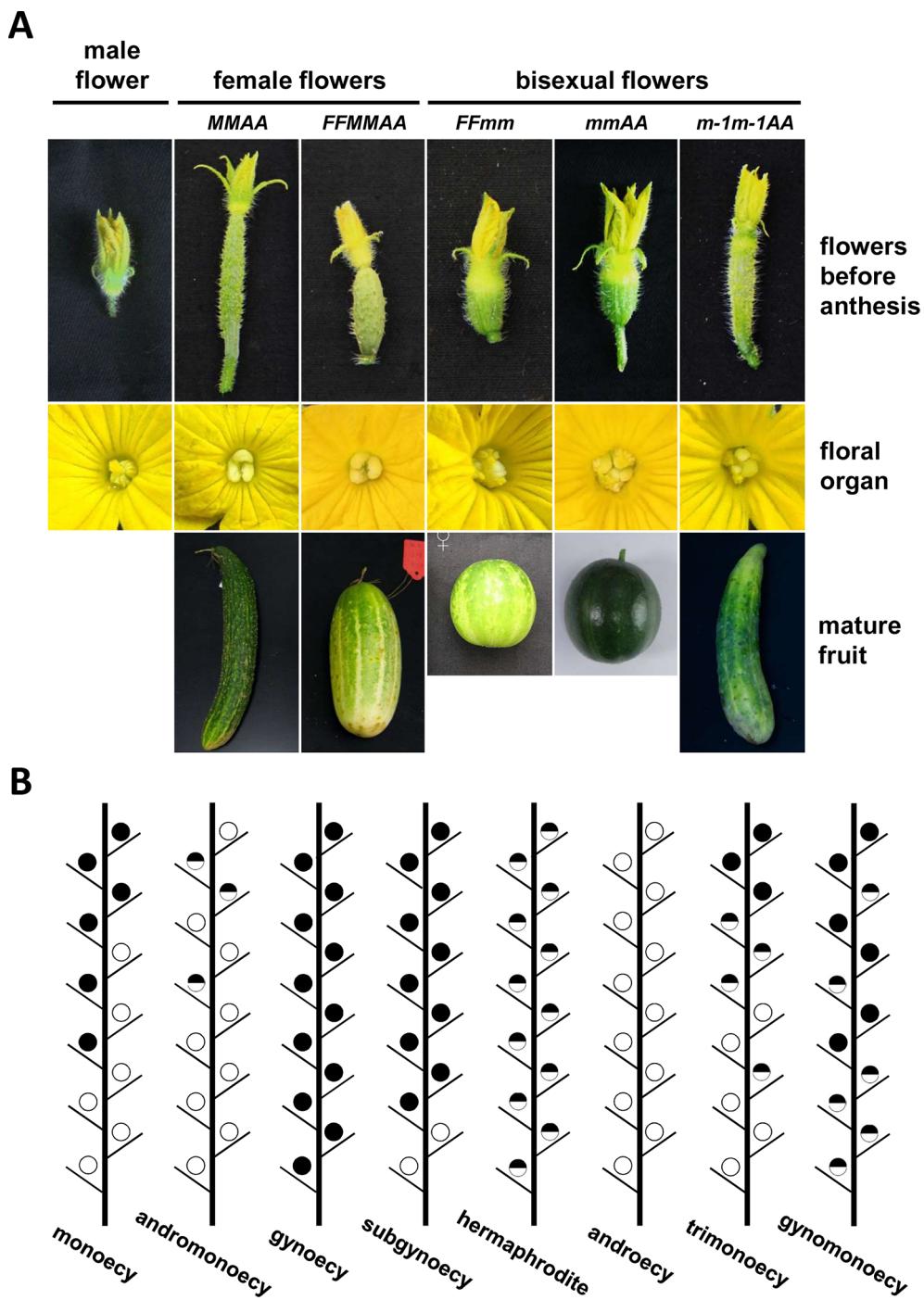
FIGURE 1 | Sex expression in cucurbits. **(A)** Flower sex type and the resulting fruit morphology in five cucumber lines with different genotypes (modified from Tan et al., 2015). **(B)** Schematic diagram of sex expression in the plant main vine in cucurbits. Black, blank, and mix circles represent female, male, and bisexual flowers, respectively. Sex type of the first blooming flower was used to defined the sex type of the node arising the flower.

sativus, Cucumis melo, Citrullus lanatus, and *Cucurbita pepo* is monoecy, in which only unisexual flowers bear. However, the distribution of male and female flowers is varied in different species or varieties. Usually, in a monoecious cucumber plant, male flowers arise in early or lower nodes, followed by a mixture of male and female flowers at the middle nodes,

and ending with female flowers only in the higher nodes. In cucumber and melon, gynoecious lines produce only female flowers, while androecious plants bear only male flowers. Male and bisexual flowers can be found in andromonoecious lines, which can be regarded as bisexual flowers replacing female flowers found in monoecious lines. Hermaphroditic plants

bear only bisexual flowers. Subgynoecious plants, which are found in some watermelon, zucchini, and cucumber lines, produce few male flowers in the beginning nodes and all female flowers in the later nodes. Their most obvious difference from the monoecious lines is the lack of the mixed phase comprising male and female flowers (Galun, 1961; Kubicki, 1969a; Kubicki, 1969b; Kubicki, 1969d). In most cases, bisexual flowers and female flowers are exclusive in a given cucumber and melon plant. However, in a cucumber mutant, certain watermelon plants, and zucchini lines treated with high temperature, hermaphroditic, female, and/or male flowers can arise in the same plant. In these cases, the sex types are named as trimonoecy (or gynomonoecy), trimonoecy, and partial andromonoecy in these three species, respectively (Kubicki, 1969c; Martínez et al., 2014; Ji et al., 2015).

In the early stage of flower development, all floral buds are morphologically hermaphrodite, containing staminate and pistillate primordia. Selective arrest of either the staminate or pistillate parts results in female or male flowers, respectively. Furthermore, if the arrest does not happen, bisexual flowers are formed (Atsmon and Galun, 1962). Results of the detailed

section assay divided the development of cucumber flowers into 12 stages (Bai et al., 2004). The floral meristem is initiated in stage 1. From stage 2 to 5, sepal, petal, staminate, and pistillate (carpel) primordia are initiated sequentially. Selective arrest then happens after stage 5. In a bud destined to be male, the stamen differentiates anther and filament in stage 6, the anther expands in stage 7, locules differentiate in stage 8, microsporocytes initiate in stage 9, meiosis initiates in stage 10, uninuclear pollen appears in stage 11, and finally, mature pollen is formed in stage 12. In male buds, from stage 6 to 12, the carpel primordia become slightly enlarged. By contrast, in a bud committed to be female, carpel primordia elongate in stage 6, carpel primordia differentiate the stigma and ovary in stage 7, the stigma elongates, and ovule and integument primordia initiate in stage 8, macrosporocytes initiate in stage 9, meiosis initiates in stage 10, embryo sac is formed in stage 11, and finally, all appendant tissues mature in stage 12. The staminate primordia in female flowers can differentiate anthers and filaments in stage 6; however, they are smaller than those in male floral buds. Thereafter, from stage 7, the arrest of stamen development is indicated by their limited size increase. Our data showed that, in bisexual flowers, staminate and pistillate
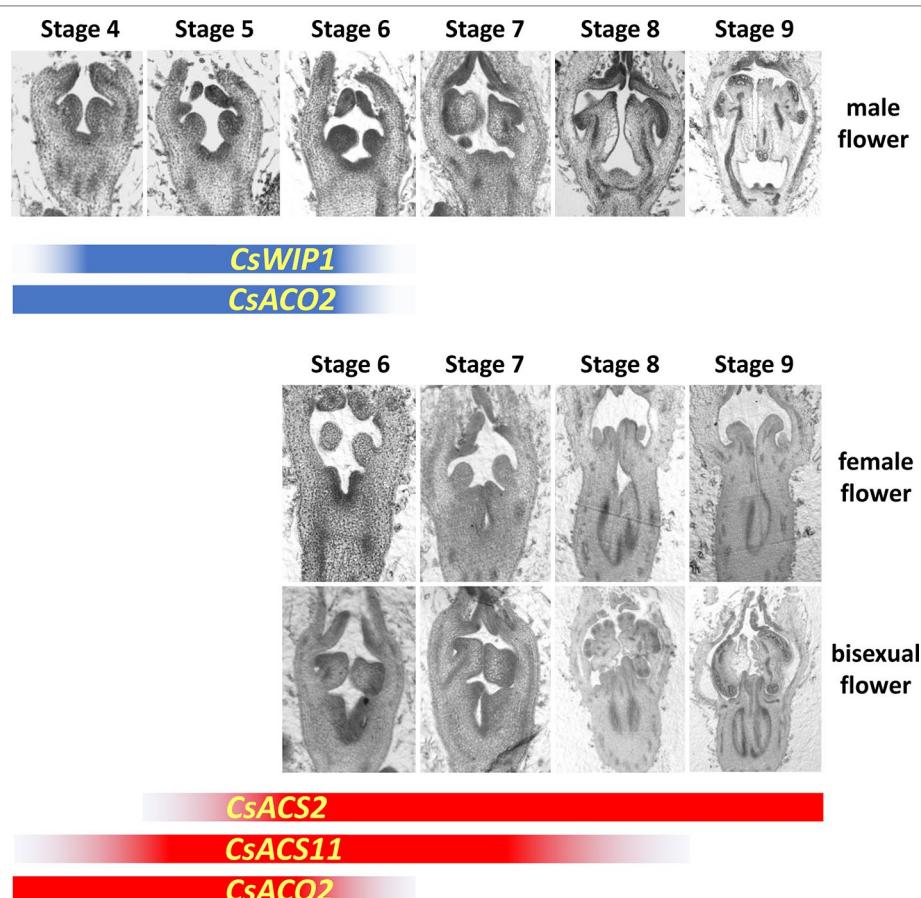
FIGURE 2 | Developmental stages of male, female, and bisexual flowers and the expression pattern of sex-controlling genes in cucumber. Before stage 5, the developmental processes in male, female, and bisexual flowers are similar under visual observation. The detailed developmental program is shown in the text. The expression durations of sex controlling genes are showed in blue (male flower) and red (female flower) bands, and the graduated dark color represents the messenger RNA (mRNA) accumulation. The section assays of male and female flowers are modified from Niu et al., 2018.

primordia show normal morphological differentiation just as in male and female buds, respectively (**Figure 2**).

Morphologically, after floral meristem initiation, the sex differentiation can be summarized as two options—pistillate initiation, which refers to the induction of pistillate primordia; simultaneous staminate and pistillate primordial growth, which is related to uni/bisexual flower development. Based on these understandings, several gene loci associated with sex expression (sex type) have been identified and cloned in the last two decades. In this review, we discuss recent advances relating to the mechanism of sex determination in the Cucurbitaceae family, beginning with the genes controlling specific processes in sex differentiation.

GENES RESULTING IN PISTILLATE PRIMORDIA INITIATION

There are two kinds of well-studied gynoecy-controlling gene loci, conferring dominant and recessive gynoecy in cucurbits. The dominant gynoecy is a unique phenotype in cucumber, which is different from other Cucurbitaceous plants. Gynoecy is a particularly important trait in cucumber breeding. Immature cucumber fruit are usually harvested a few days after flowering. Therefore, combined with parthenocarpy, more female flowers mean higher yield in cucumber production. In 1935, Tkachenko described the first gynoecious cucumber in a Japanese (Korean) variety (Kubicki, 1969a). Because there are no male flowers on the gynoecious plants, homozygous gynoecious varieties were nonhereditary before the 1960s, when gibberellic acid (GA) was first used to induce male flowers (Peterson and Anhder, 1960). In 1961, Shifriss explained that a gene (named as *Acr*) could act as an accelerator to impel female flowers to lower nodes (Shifriss, 1961). Later, Galun (1961) and Kubicki (1969a) named the locus as st^F and Acr^F, respectively. In 1976, the symbol *F* (*Female*) was finally confirmed to represent the dominant gynoecy-controlling locus (Robinson et al., 1976). Later studies indicated that the *F* locus was associated with an additional copy (*CsACS1G*, G = gynoecy) of a 1-aminocyclopropane-1-carboxylic acid synthase gene, *CsACS1* (Trebitsh et al., 1997). The open reading frame and proximal promoter (−410 bp upstream sequence) are almost identical between the two genes. The different distal promoter sequence of *CsACS1G* is homologous to that of a putative *branched-chain amino acid transaminase* (*BCAT*) gene (Mibus and Tatlioglu, 2004; Knopf and Trebitsh, 2006). Bioinformatic analysis discovered a copy number variant, which arose from a 30-kb genomic sequence duplication (including *CsACS1* and *BCAT*), involving in the *F* locus. The copy number variant region might present a tandem repeat of the original 30-kb region in gynoecious lines, and the "junction point" of the two repeats is *CsACS1G* (Zhang et al., 2015).

Natural melon and watermelon varieties possess recessive gynoecy loci, which are named as *g* (*gynoecious*) and *gy* (*gynoecious*), respectively. Poole and Grimball (1939) reported a recessive *g* locus that controls gynoecy or subgynoecy in melon. Martin et al. (2009) identified the *g* gene as *CmWIP1*, encoding a C2H2 zinc-finger-type transcription factor.

Expression of *CmWIP1* leads to carpel abortion, resulting in male flowers. CmWIP1 indirectly represses the expression of *CmACS7*, which is the andromonoecious gene introduced later. In gynoecious lines, an insertion of a transposon (1.3 kb downstream of the gene) represses the expression of *CmWIP1 via* epigenetic changes in its gene promoter. In watermelon, a chromosome translocation produced an insertion mutation in the *ClWIP1* gene (the *CmWIP1* ortholog), leading to a gynoecious line (Zhang et al., 2019). Natural mutants of *CsWIP1* in cucumber varieties are unavailable. Using the clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 technology, the created *CsWIP1*-editing mutant lines also showed gynoecy (Hu et al., 2017). All these studies confirmed that WIP1 is a conserved regulator of sex determination in cucurbits.

It should be noted that all the genes controlling gynoecy described above are associated with carpel-bearing flowers. Therefore, the genes function not only in female flowers but also in bisexual flowers. The gynoecy-controlling genes induce (or release) pistillate initiation in female and hermaphroditic flowers.

MUTANT GENES RESULTING IN THE HERMAPHRODITIC PHENOTYPE

Although monoecy is the most common sex type in cucurbits, melon breeders prefer bisexual flowers, making andromonoecy predominant in melon compared with that in other cucurbits (Boualem et al., 2008). Bisexual flowers make natural pollination easier, and the resulting fruits are usually rounder than the products of female flowers (Li et al., 2009; Aguado et al., 2018), both of which are desired phenotypes for melon production. An interesting phenomenon in bisexual flowers is that the fertilization and seed-setting ability in melon is much higher than that in cucumber, which might be the result of long-term domestication in melon breeding.

In cucurbits, the genes controlling the hermaphroditic phenotype are highly conserved. Rosa (1928) stated that andromonoecy is a recessive character in *Cucumis* and *Citrullus*. In melon, the gene locus was named as *a* (*andromonoecious*) before 2015, then changed to *m* (*monoecious*) to avoid confusion with the androecy controlling gene (*a*). A single-nucleotide mutation in *CmACS7* was identified as associated with andromonoecy in melon. *CmACS7* also encodes a 1-aminocyclopropane-1-carboxylic acid synthase like *CsACS1G*, and the mutation severely loses the enzymatic activity. Expression of *CmACS7* inhibits staminate development in female flowers but is not required for carpel development (Boualem et al., 2008).

Two kinds of natural *monoecious* mutations (named *m* and *m-1*) in cucumber have been identified, and both of which are associated with *CsACS2*, which is an ortholog of *CmACS7* (Boualem et al., 2009; Li et al., 2009; Tan et al., 2015). The mutation in the *m* allele is also a single-nucleotide change; however, it mutates another conserved active site residue, different from the mutation in melon. In addition, a 14-bp deletion is found in the third exon of *CsACS2* in the *m-1* allele, which deduces a truncated protein. The mutations result in severe loss of enzyme

activity in plants with the *m* allele and total loss in those with the *m-1* allele.

The *CmACS7* orthologs, *CitACS4*/*ClACS7* and *CpACS27A*, are associated with andromonoecy in watermelon and zucchini, respectively (Martínez et al., 2014; Boualem et al., 2016; Ji et al., 2016; Manzano et al., 2016). In watermelon, the isoforms encoded by *ClACS7* in andromonoecious lines showed no enzymatic activity, and the isoform in the monoecious line was active. In zucchini, even though neither of the parental lines showed a standard andromonoecious phenotype, a mutant nucleotide in *CpACS27A* was considered to be necessary, but not sufficient, to confer partial andromonoecy.

Pleiotropy is another characteristic of the genes controlling the hermaphroditic phenotype. Usually, bisexual flowers produce rounder fruits than female flowers (**Figure 1A**). An interesting finding in cucumber was that the trait of spherical fruit cosegregated with the *m* allele in two large F_2 populations comprising 5,500 individuals in total (Li et al., 2009). Recently, a study of cucumber fruit growth confirmed that *CsACS2* participates in fruit elongation *via* regulation of ubiquitination (Xin et al., 2019), which supplies a link between sex type and fruit development. In addition, hermaphroditic mutations affect floral organ development. In watermelon, the mutant allele cosegregated with slower growth and maturation of petals and carpels, which resulted in delayed anthesis time in hermaphrodite flowers. Moreover, the number of fruit and seed set were lower in the mutant lines, representing reduced fertilization activity of bisexual flowers in watermelon like in cucumber (Aguado et al., 2018). In zucchini, the bisexual flowers also showed delayed development and maturation of petals and a higher ovarian growth rate (Martínez et al., 2014).

Summarizing the current findings, in cucurbits, the *ACS7* orthologs are expressed in pistil-bearing flowers but not in male flowers. The expression of functional isoforms arrests staminate primordia, producing female flowers, while the nonfunctional or mutant isoforms lose this function and allow staminate and pistillate primordia to grow simultaneously, resulting in bisexual flowers. Therefore, in addition to mutation research, analyzing the regulation of gene expression is also important for the *ACS7* orthologs.

MUTANT GENES LEADING TO ANDROECY

In cucumber and melon, sex expression in the main vine and lateral branches are usually different. Usually, the first several nodes in the lateral branch have high feminization potential, producing female flowers in monoecious lines and bisexual flowers in andromonoecious lines. Strictly, a variety without any pistil-bearing flowers in the main vine and lateral branches is defined as androecy; otherwise, it is identified as monoecy (with female flowers) or andromonoecy (with bisexual flowers). Obviously, an androecious line has little economic value in production, and the existing varieties are all mutants. In cucumber, a recessive *a* locus was identified to intensify the androecious nature (Kubicki, 1969b). The gene is hypostatic to the *F* gene, and a plant with a genotype of *ffaa* is completely male.

A rare androecious cucumber variety "EREZ" helped to clone the *a* gene, for which the wild-type allele is *CsACS11*, encoding the third 1-aminocyclopropane-1-carboxylic acid synthase involved in sex determination. Similar to the *F* and *M* genes, the mutant isoform isolated from "EREZ" had no enzymatic activity (Boualem et al., 2015). Using a targeting-induced local lesions in genomes strategy, 10 mutations in the melon ortholog of *CmACS11* were created, and two lines containing changes in highly conserved amino acids were observed as androecious.

Besides the traditional *a* locus, an ethyl methanesulfonate-induced mutation helped to discover the second androecious cucumber variety, and the mutation was identified in *CsACO2*, encoding 1-aminocyclopropane-1-carboxylic acid oxidase. The single-nucleotide change in the mutant gene resulted in an inactivated enzyme (Chen et al., 2016). The melon orthologous gene of *CsACO2* is *CmACO3*. Both of these genes showed similar expression patterns (see below).

OTHER SEX-TYPE RELATED LOCI IDENTIFIED *VIA* GENETIC ANALYSIS

Previous genetic studies have identified many loci that control conventional and accidental sex mutations. In cucumber, besides *F*, *m*, and *a*, the Intensive Female (*In-F*) gene was identified as increasing the female flower ratio in monoecious plants (without the *F* gene). In addition, a plant with both *F* and *In-F* genes could not produce male flowers when treated with GA (Kubicki, 1969b). Kubicki (1969b) also described an accelerator gene (*acr¹*), conferring continuous nodes with female flowers in monoecious lines. In subgynoecious cucumber lines, a consistent major quantitative trait locus, which mainly increased the degree of femaleness, was identified on chromosome 3 (*sg3.1*) from two independent studies (Ji et al., 2016; Win et al., 2019). The relationship between *acr¹* or *In-F* and the genes should be clarified in future studies. Using artificial mutagenesis, Kubicki (1974, 1980) identified the hermaphrodite (*h*) and gynoecious (*gy*) loci. Unlike the *m* gene, the *h* gene governs bisexual flowers with normal ovaries as in female flowers, including their shape and pollination ability. However, analysis of the data did not allow us to determine the relationship or difference between the *h* gene and the *m-1* mutation. The recessive *gy* gene was described as intensifying femaleness in cucumber and is linked with the *F* gene. The function of *gy* gene was similar to the *g* gene in melon. However, *CsWIP1* (the melon *g* ortholog in cucumber) resides on chromosome 4, and the *F* gene is on chromosome 6, which means that *gy* and *g* might be two different genes. Trimonoecious plants have been reported in cucumber, watermelon, and zucchini (Kubicki, 1969c; Martínez et al., 2014; Ji et al., 2015). In cucumber, the gene responsible was named as *tr* (trimonoecious), while the phenomenon is controlled by the *tm* gene in watermelon (to date, no name has been given in zucchini). However, the detailed structures of bisexual flowers in the three species are different. In cucumber, the bisexual flowers occurring in trimonoecious plants have superior ovaries (hypogynous, the normal bisexual and female flowers are epigynous), derived as a modification of staminate flowers, while the bisexual flowers

in trimonoecious watermelon and zucchini seem to be same as those in andromonoecious plants. Unfortunately, standard plant materials possessing the above loci (*In-F*, *acr¹*, *h*, *gy*, *tr*) are not widespread. We look forward to seeing in-depth studies and cloning of these genes in the future.

ETHYLENE AND SEX DETERMINATION IN CUCURBITS

The most important factor regulating sex expression in cucurbits is the phytohormone ethylene, which controls the transition of female flowering and the ratio of female flowers (Byers et al., 1972; Atsmon and Tabbak, 1979; Owens et al., 1980; Takahashi et al., 1982; Takahashi and Jaffe, 1984; Kamachi et al., 1997; Trebitsh et al., 1997). In cucumber and melon, ethylene (or its releasing agent) has been used to induce female flowers for decades (Rudich et al., 1969; Tsao, 1988; Yin and Quinn, 1995). In zucchini, sex determination in individual floral bud appears to be regulated by ethylene in a similar way (Manzano et al., 2010a; Manzano et al., 2010b; Manzano et al., 2011; Manzano et al., 2013). By contrast, inhibition of ethylene biosynthesis or perception leads to increased maleness in cucumber, melon, and zucchini (Byers et al., 1972; Owens et al., 1980; Manzano et al., 2011). The relationship between ethylene and the sex type in watermelon is complex. In watermelon, female flowers require much more ethylene than male flowers to develop. In addition, bisexual flowers result from a decrease in ethylene production in female floral buds, and ethylene is required to arrest the development of stamens in female flowering, similar to the process in cucumber and melon. Nevertheless, ethylene inhibits the transition from male to female flowering and reduces the number of pistillate flowers, which contrasts with the findings in other cucurbits (Manzano et al., 2014; Zhang et al., 2017a). An interesting phenomenon was observed in watermelon, in which ethephon (an ethylene-releasing reagent) treatment induced numerous abnormal flowers in gynoecious and hermaphroditic plants (Zhang et al., 2017a).

In cucumber and melon, different ethylene responses in staminate and pistillate primordia are used to explain the selective arrest occurring during sex determination. It has been proposed that differing levels of sensitivity in the stamen or carpel primordia could allow each type of primordium to react independently to different ranges of ethylene concentrations (Yin and Quinn, 1995). A higher ethylene threshold for stamen suppression than carpel promotion, coupled with the timing of the increase in ethylene production occurring after the carpels are established, would prevent stamen inhibition before carpel establishment, thereby ensuring the development of flowers (Switzenberg et al., 2014). Ectopic expression of ethylene-related genes suggested that ethylene perception by stamen primordia, but not carpel primordia, is essential for the production of carpel-bearing buds (Little et al., 2007; Switzenberg et al., 2014). Ethylene might promote female flower development *via* an organ-specific induction of DNA damage in primordial anthers. The organ-specific ethylene perception might require downregulation of *CsETR1* (encoding an ethylene receptor protein, see below)

expression and increased expression of *CsCaN* (encoding a calcium-dependent nuclease) (Wang et al., 2010; Gu et al., 2011).

Ethylene synthesis results from the activity of 1-aminocyclopropane-1-carboxylic acid (ACC) synthase (ACS) and 1-aminocyclopropane-1-carboxylic acid oxidase (ACO), which transform S-adenosyl-L-Met (SAM) into ACC and convert ACC into ethylene, respectively (Adams and Yang, 1979; Yang and Hoffman, 1984). After biosynthesis, ethylene signaling is perceived by the receptor proteins, which are located in the endoplasmic reticulum. The receptors are negative regulators of ethylene signaling, and in the absence of ethylene, the receptors activate constitutive triple-response 1 (CTR1), which suppresses the ethylene response *via* inactivation of ethylene insensitive 2 (EIN2). Ethylene binding to the receptors switches off the CTR1 phosphorylation activity and activates EIN2. The C terminus of EIN2 is cut and moves into nucleus, stabilizing ethylene insensitive 3/EIN3-like (EIN3/EIL) transcription factors, which can activate the expression of target genes, including those encoding ethylene response factor (ERF) transcription factors. The ERFs then initiate the expression of downstream ethylene-responsive genes (Zhang et al., 2009; Klee and Giovannoni, 2011; Liu et al., 2015).

To date, except for *WIP1* orthologous genes, other sex-controlling genes, including *CsACS1G*, *CsACS2*, *CsACS11*, and *CsACO2* in cucumber, *CmACS7* and *CmACS11* in melon, *CitACS4/ClACS7* in watermelon, and *CpACS27A* in zucchini, have important roles in ethylene biosynthesis. Because ethylene participates in sex determination directly in cucurbits, the identification of many sex-related ethylene synthases is not surprising. However, since nearly all the genes show similar biochemical function (producing ethylene), the regulation of their expression should be important. Moreover, a high concentration of ethylene is harmful to young tissue, which was observed in cucumber protoplasts and watermelon plants (Wang et al., 2010; Zhang et al., 2017a). Therefore, the specific spatiotemporal and coordinate expression of the *ACS* and *ACO* genes, which produce local ethylene accumulation, inducing pistil and arresting stamen development, is critical in sex determination.

TRANSCRIPTIONAL CHARACTERISTICS OF THE SEX-RELATED GENES IN SEX DETERMINATION

Studies with exogenous ethylene have indicated that the timing and concentration are key factors that determine whether carpel or stamen development is affected (Switzenberg et al., 2014). Therefore, the spatiotemporal expression pattern of the sex-related genes should be studied. The developmental process of flower buds reveals that sex determination happens between stage 5 and 6; therefore, all the regulatory genes should function before, or at least no later than, these two periods. *CsACS1G*, *CsACS11*, *CsACS2*, *CmACS11*, *CmACS7*, and *ClACS7/CitACS4* are only expressed in female flowers, while *CsWIP1*, *CmWIP1*, and *ClWIP1* are expressed in male flowers. The transcription of *CsACO2* and *CmACO3* has no sex specificity. *CsACS1G* was considered to be autonomously expressed in the shoot

or early flower bud before all the other sex-controlling genes in gynoecious cucumber (Knopf and Trebitsh, 2006; Li et al., 2012). However, its detailed expression pattern is still unknown because its low accumulated messenger RNA level limits the use of *in situ* hybridization assays. Transcripts of all the bisexual flower controlling genes, *CsACS2*, *CmACS7*, and *ClACS7/CitACS4*, began to accumulate just beneath the pistil primordia of flower buds from stage 5, and then continued to accumulate in central region of the developing ovary (Saito et al., 2007; Boualem et al., 2008; Boualem et al., 2016). The expression signals of both *CsACS11* and *CmACS11* were first detected below the carpel primordia from stage 4 and continued at least until stage 8 (Boualem et al., 2015). *CsACO2* and *CmACO3* expression was first detected in the center of stage 2 to 4 flower buds, just beneath the location of future carpel primordia, and remained expressing in the carpel and stamen at a relatively low levels after stage 6 (Chen et al., 2016). In male flowers, although *CmWIP1* seems to have an enhanced expression compared with *CsWIP1*, they are both expressed from stage 4 to 6 (Boualem et al., 2015; Chen et al., 2016). The expression pattern of sex-controlling genes is summarized in **Figure 2**. The interacting order of these sex-controlling genes in sex determination can be deduced from the sequence and duration of their gene expression.

Ethylene is also a key regulator of the expression of sex-controlling genes. Treatment with exogenous ethylene at an appropriate concentration increased the transcription of *CsACS1*, *CsACS2*, *CsACS11*, *CmACS11*, and *CmACS7* and downregulated that of *CsWIP1* and *CmWIP1* (Yamasaki et al., 2001; Li et al., 2012; Switzenberg et al., 2014; Tao et al., 2018). Endogenous ethylene produced by a first expressed sex-specific gene might also act with other sex-controlling genes, which is used to explain the interacting phenomenon among them (see below). A hypothesis was proposed that ethylene mediated the

interaction among different sex-controlling genes, including (to date, at least) *CsACS2*, *CsACS11*, *CmACS11*, *CmACS7*, *CsWIP1*, and *CmWIP1*. Cloning of ethylene signaling factors, CsERF110/CmERF110 and CsERF31, which directly combine with the promoters of *CsACS11/CmACS11* and *CsACS2* to activate their expression, respectively, supplied evidence for this hypothesis (Pan et al., 2018; Tao et al., 2018).

Other hormones, like auxin, brassinosteroids (BRs), and GA, are also involved in floral sexual differentiation, all of which might function *via* influencing ethylene biosynthesis or signaling (Rudich et al., 1972; Trebitsh et al., 1987; Yin and Quinn, 1995; Papadopoulou and Grumet, 2005; Zhang et al., 2014a). Auxin increases the expression of *ACS* genes, inducing female flowers (Trebitsh et al., 1987). BRs also increase ethylene production and indirectly participate in cucumber sex determination, in which CsPSTK1 (a putative serine/threonine kinase) might be involved (Papadopoulou and Grumet, 2005; Pawełkowicz et al., 2012). Exogenous GA inhibits ethylene biosynthesis, in which CsGAMYB1 (a GAMYB homolog, positive regulator of GA signaling pathway) and CsGAIP (a DELLA homolog) might have functions (Zhang et al., 2014a; Zhang et al., 2014b).

GENE INTERACTION CONFERRING SEX EXPRESSION

Classical genetic analyses helped to propose a systematic phenotype–genotype relationship for each sex type in cucumber, melon, and watermelon (Poole and Grimball, 1939; Galun, 1961; Kubicki, 1969a, Kubicki, 1969b; Kubicki, 1969d; Robinson et al., 1976; Kenigsbuch and Cohen, 1987; Kenigsbuch and Cohen, 1990; Ji et al., 2015). Here, we tried to integrate the results from genetic studies, biochemical assays, and physiological responses (**Figure 3**).
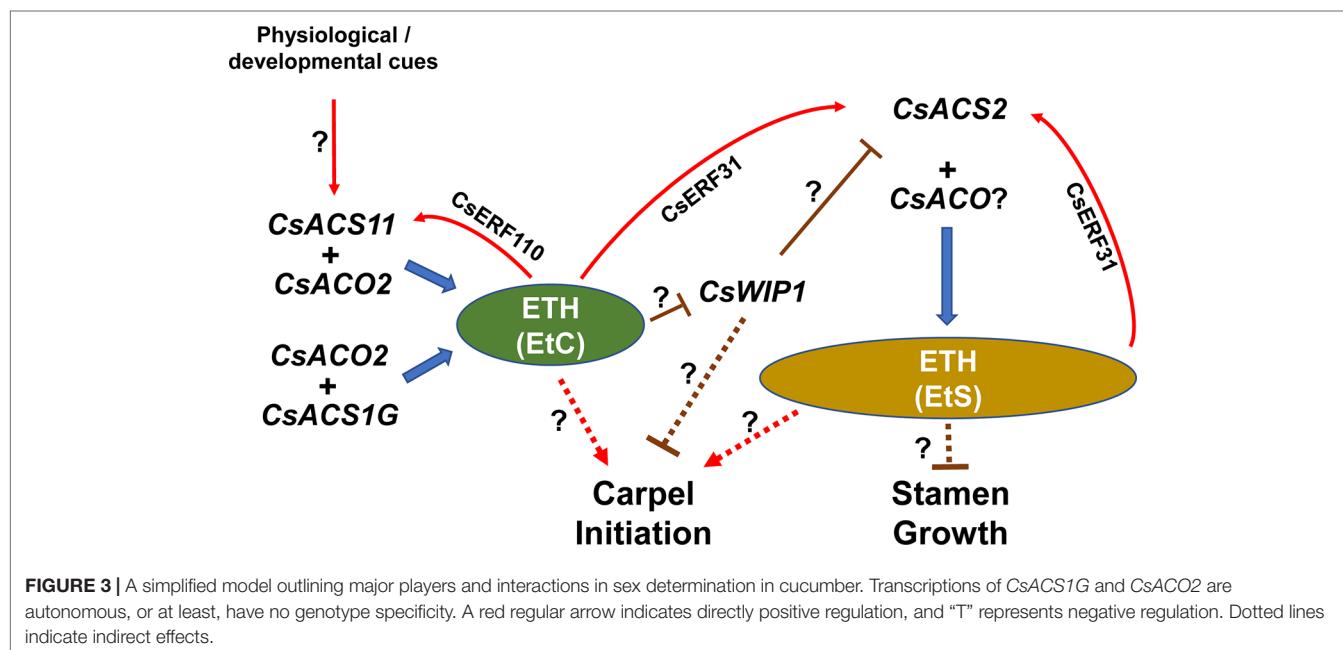
FIGURE 3 | A simplified model outlining major players and interactions in sex determination in cucumber. Transcriptions of *CsACS1G* and *CsACO2* are autonomous, or at least, have no genotype specificity. A red regular arrow indicates directly positive regulation, and "T" represents negative regulation. Dotted lines indicate indirect effects.

In cucumber, because there are different levels of sensitivity to ethylene in the stamen and carpel, two ethylene thresholds are proposed to define for carpel promotion (EtC) and stamen suppression (EtS), and EtS is believed to be higher than EtC (Switzenberg et al., 2014). Thus, the genotype–phenotype relationship is proposed as:

(1) The *F* gene (*CsACS1G*) is autonomously expressed, producing ethylene to reach the EtC (but not higher than the EtS), which initiates pistillate primordia. The ethylene also can induce *M* gene (*CsACS2*) expression, synthesizing higher (and/or long-term) ethylene content reaching the EtS, and arresting staminate primordia. Consequently, a combination of *F* and *M* genes (*FFMM*) produces continuous female flowers and confers the gynoecious phenotype;

(2) When the genotype is *FFmm*, EtC can be achieved by *CsACS1G* expression, and the pistil develops normally. However, the mutant *m* gene encodes an inactive ACS, leading to insufficient ethylene to reach the EtS; thus, the stamen develops. Therefore, the *FFmm* genotype results in a plant with all nodes bearing bisexual flowers, producing a hermaphroditic line;

(3) When the *F* locus is homozygous recessive (*ff*), the plant sex type is dependent on the expression pattern of the *A* gene (*CsACS11*). If *CsACS11* is expressed, the ethylene becomes higher than the EtC, but lower than the EtS, which can induce pistil initiation and activate *CsACS2* expression, finally producing female flowers. If *CsACS11* is silent, there is insufficient ethylene (< EtC); therefore, the pistillate primordia cannot initiate, while the stamen development is released, resulting in male flowers. Combining these two conditions, the genotype *ffMMAA* results in a monoecious individual;

(4) Like the phenomenon in hermaphroditic lines, when the *m* locus is homozygous recessive, stamens in female flowers are not suppressed, and genotype *ffmmAA* results in andromonoecious plants;

(5) When both of *F* and *A* loci are mutated (*ffaa*), no genes can produce ethylene up the level of the EtC, in which case *CsACS2* and pistillate primordia cannot be induced. Therefore, only male flowers can develop, and an androecious plant emerges (Li et al., 2012; Tao et al., 2018).

Because no natural *CsWIP1* and *CsACO2* mutations were used in these previous genetic analyses, all the plants studied were assumed to have wild-type *CsWIP1* and *CsACO2* genes. Analysis in melon showed that *CmWIP1* negatively regulates the expression of *CmACS7* (its ortholog in cucumber is *CsACS2*) (Martin et al., 2009). Ethylene could downregulate the expression of *CsWIP1*, meaning that the EtC may induce *CsACS2 via* suppressing *CsWIP1*. In a mutant *Cswip1* background, the expression of *CsACS2* is released, producing the EtS directly, which is enough to initiate pistillate primordia and arrest staminate primordia, resulting in female flowers and gynoecious plants (Hu et al., 2017). In addition, CsWIP1 might directly suppress pistillate primordia initiation *via* an unknown pathway, which was proposed in melon (Boualem et al., 2015). The enzyme encoded by *CsACO2* is considered to at least combine with *CsACS1G* and *CsACS11* to complete ethylene synthesis. Therefore, *Csaco2* mutants break the formation of EtC, resulting in an androecious phenotype

(Chen et al., 2016). In the future, the sex type of the *Cswip1Csaco2* double mutant should be investigated to identify the relationship between *CsACS2* and *CsACO2* in cucumber.

In melon, except for the dominant *F* gene, all the sex-type-related genotype–phenotype relationships are similar to those in cucumber. Therefore, plant femaleness is dependent on mutant *Cmwip1* (the *g* gene in melon) and/or *CmACS11* expressing (the *A* gene in melon). The bisexual flowers in melon are the result of mutations in *CmACS7* (the *m* gene in melon). Classical genetics confirmed that the genotype *MMAAGG* results in a monoecious plant, *mmAAGG* results in andromonoecy, *MMgg* results in gynoecy, *mmgg* results in a hermaphrodite, and *aaGG* results in androecy (Poole and Grimball, 1939; Kenigsbuch and Cohen, 1987; Kenigsbuch and Cohen, 1990). An interestingly different phenomenon was observed between monoecious cucumber and melon plants with the same (*ff*)*MMAA*(*GG*) genotype that femaleness on the main vine of cucumber (although it is often changeable) is higher than that in melon (usually all nodes produce male flowers). This might refer to different expression activity of the *A* genes in these two species. We have identified that ethylene could induce the expression of *CsACS11* (Tao et al., 2018). However, there is no *CsACS1G* in monoecious lines to autonomously produce ethylene. Therefore, other physiological or developmental cues inducing *CsACS11* or *CmACS11* should be identified in future studies, which might help to explain the problem stated by Ma and Pannell (2016) that "what decides whether *ACS11* is on or off in particular flowers."

In watermelon, three recessive alleles were suggested to control the sex types: <u>a</u>ndromonoecious (*a*), <u>gy</u>noecious (*gy*), and <u>tri</u>monoecious (*tm*) (Ji et al., 2015). Therefore, phenotype–genotype relationships are proposed as: monoecious, *AAGyGyTmTm*; trimonoecious, *AAGyGytmtm*; andromonoecious, *aaGyGy*; gynoecious, *AAgygyTmTm*; gynomonoecious, *AAgygytmtm*; and hermaphroditic, *aagygy*. Recent gene cloning helped to identify that the *a* gene is a mutation in *ClACS7/CitACS4*, and the *gy* gene is mutated in *ClWIP1*, which are orthologous genes of *CsACS2/ CmACS7* and *CsWIP1/CmWIP1*, respectively (Boualem et al., 2016; Ji et al., 2016; Manzano et al., 2016; Zhang et al., 2019). Therefore, it seems that the *WIP1-CsACS2/CmACS7* relationship is conserved in all the studied cucurbits.

There is little information about the phenotype–genotype relationship in zucchini. Sex determination in individual floral buds of zucchini appears to be regulated by ethylene in the same way as that in melon and cucumber, and the *ACS7* ortholog *CpACS27A* is also involved in bisexual flower development (Manzano et al., 2010a; Manzano et al., 2010b; Manzano et al., 2011; Martínez et al., 2014). Therefore, we believe that the conserved *ACS11-WIP1-CsACS2/CmACS7/CpACS27A* pathway also exists in zucchini.

SUGGESTION FOR GENE NOMENCLATURE

An imminent work is unifying gene names controlling the similar sex types in cucurbits. For example, the abbreviated names of *andromonoecious* (which is now *m*, but used to be *a* before 2015) and *androecious* (is now *a*) in melon, easily cause confusion. Moreover, the genotype symbols in watermelon are also liable

to cause misunderstanding. Because all the genes controlling the appearance of bisexual flowers are *Arabidopsis ACS7* orthologs, we suggest that the symbol *m* is used for the andromonoecious phenotype, just as in cucumber and melon. Likely, the symbol *g* is suggested to use for the recessive gynoecy produced by *wip1* ortholog mutation. The structures of bisexual flowers in cucumber trimonoecious mutant (hypogynous) and in trimonoecious watermelon (seemingly normal) are different. Therefore, it is necessary to retain the current gene nomenclature (*tr* and *tm*).

Another gene symbol that needs to be discussed is *f* in cucumber. When we reexamined the *F* gene and its genomic structure, it was clear that the *F* locus is unique in gynoecious cucumber lines with a tandem 30-kb repeat (Zhang et al., 2015). This locus does not exist in lines with only one copy of the 30-kb region, as in monoecious, andromonoecious, and androecious lines. Previously, genotypes of these latter three lines were usually written as homozygous *ff* for this locus, which is used to represent the recessive allele in the *F* locus. However, we now know that there is only one form of gene (*CsACS1G*) in this locus, and no studies have demonstrated a mutant or a nonfunctional allele. In traditional cognition, the dominant *F* gene is *CsACS1G*, and the recessive *f* is considered as *CsACS1*. However, this is not right. All cucumber lines tested have *CsACS1*, and only gynoecious plants possess both *CsACS1* and *CsACS1G*. These findings mean that *CsACS1* and *CsACS1G* are not alleles of the same gene, which are not located in a same locus (30 kb apart approximately). Therefore, at this time, before a *CsACS1G* mutation is discovered, we suggest that the nomenclature of the *f* gene makes no sense and should be omitted in the genotype.

Consequently, we suggest the phenotype–genotype relationships in cucurbits as: monoecious, *MMAAGG*; andromonoecious, *mmAAGG*; androecious, *aaGG*; gynoecious, *FFMM* (in cucumber only) or *MMgg*; hermaphroditic, *FFmm* (in cucumber only) or *mmgg*; trimonoecious, *MMAAGGtrtr* in cucumber and *MMGGtmtm* in watermelon; gynomonoecious, *FFMMtrtr* in cucumber and *MMggtmtm* in watermelon.

OTHER ASPECTS RELATING TO SEX TYPE

Detailed descriptions about many of the transcriptomic, epigenomic, and metabolomic research related to sex type are beyond the scope of this manuscript (Miao et al., 2011; Wang et al., 2014; Gao et al., 2015; Zhang et al., 2017b; Lai et al., 2017; Latrasse et al., 2017; Lai et al., 2018a; Lai et al., 2018b; Song et al., 2018; Wang et al., 2018; Zhou et al., 2018; Wang et al., 2019; Pawełkowicz et al., 2019b). The sex-related genes and cues identified in these studies are associated with temperature, photoperiod, blue/red light, hormone synthesis and signaling, lipid and sugar metabolism, the cell cycle, etc. However, we do not know whether the genes are causes or results of the sex-type changes, and we cannot summarize the accurate locations of these genes in the gene pathway of sex determination.

GENERAL CHARACTERISTICS OF SEX CONTROLLING GENES

It is not surprising that near all the sex controlling genes are "ethylene synthases." Therefore, the expression regulation of

each gene should be conducted, and the specific transcription regulators for a given sex-control gene should be identified. In addition, exploring new sex-related mutations has always been a priority in sex determination research. Considering all the known genes that directly control the sex types in cucurbits, we propose that a sex-related gene may have more than one of the following characteristics: (1) the gene product directly participates in ethylene synthesis or signal transduction, (2) the gene or its product directly or indirectly regulates a known sex-control gene, (3) gene expression responds to ethylene or the factors interfering with ethylene synthesis or signaling, and (4) critically, mutation of the gene can change sex type. We believe that these characteristics could help to identify new sex-related genes in the future.

FUTURE PROSPECTS

The mechanism of sex determination is of great interest to researchers. Meanwhile, the close relationship between sex type and yield in cucurbits has attracted increased attention in plant breeding. At present, a model of the ethylene core has been established in four cucurbit species (cucumber, melon, watermelon, and zucchini). However, the direct regulators and the molecular details remain poorly understood. Exploring more mutations and using reverse genetics are the most effective way to identify a gene-controlling sex differentiation. In addition, since the critical developmental stage in sex determination is clear, more precise approaches, such as laser microdissection and single-cell RNA sequencing have the potential to reveal the detailed gene pathways involved in this process. We hope that the suggestions proposed in this review are conducive to revealing the mechanisms of sex determination in cucurbits.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adams, D. O., and Yang, S. F. (1979). Ethylene biosynthesis: identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. *Proc. Natl. Acad. Sci. USA* 76, 170–174. doi: 10.1073/pnas.76.1.170

Aguado, E., García, A., Manzano, S., Valenzuela, J. L., Cuevas, J., Pinillos, V., et al. (2018). The sex-determining gene *CitACS4* is a pleiotropic regulator of flower and fruit development in watermelon (*Citrullus lanatus*). *Plant Reprod.* 31, 411–426. doi: 10.1007/s00497-018-0346-1

Atsmon, D., and Galun, E. (1962). Physiology of sex in *Cucumis sativus* L. Leaf age patterns and sexual differentiation of floral buds. *Annal. Bot.* 26, 137–146. doi: 10.1093/oxfordjournals.aob.a083781

Atsmon, D., and Tabbak, C. (1979). Comparative effects of gibberellin, silver nitrate and aminoethoxyvinylglycine on sexual tendency and ethylene evolution in the cucumber plant (*Cucumis sativus* L.). *Plant Cell Physiol.* 20, 1547–1555. doi: 10.1093/oxfordjournals.pcp.a075957

Bai, S., Peng, Y., Cui, J., Gu, H., Xu, L., Li, Y., et al. (2004). Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus L.*). *Planta* 220, 230–240. doi: 10.1007/s00425-004-1342-2

Bhowmick, B. K., and Jha, S. (2015). Dynamics of sex expression and chromosome diversity in Cucurbitaceae: a story in the making. *J. Genet.* 94, 793–808. doi: 10.1007/s12041-015-0562-5

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., et al. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* 321, 836–838. doi: 10.1126/science.1159023

Boualem, A., Lemhemdi, A., Sari, M-A., Pignoly, S., Troadec, C., Abou Choucha, F., et al. (2016). The andromonoecious sex determination gene predates the separation of *Cucumis* and *Citrullus* Genera. *PLoS ONE* 11, e0155444. doi:10.1371/journal.pone.0155444

Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M-A., et al. (2015). A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* 350, 688–691. doi: 10.1126/science.aac8370

Boualem, A., Troadec, C., Kovalski, I., Sari, M.-A., Perl-Treves, R., and Bendahmane, A. (2009). A conserved ethylene biosynthesis enzyme leads to andromonoecy in two Cucumis species. *PLoS ONE* 4, e6144. doi:10.1371/journal.pone.0006144

Bu, F., Chen, H., Shi, Q., Zhou, Q., Gao, D., Zhang, Z., et al. (2015). A major quantitative trait locus conferring subgynoecy in cucumber. *Theor. Appl. Genet.* 129, 97–104. doi: 10.1007/s00122-015-2612-z

Byers, R. E., Baker, L. R., Sell, H. M., Herner, R. C., and Dilley, D. R. (1972). Ethylene: a natural regulator of sex expression of *Cucumis melo* L. *Proc. Natl. Acad. Sci. USA.* 69, 717–720. doi: 10.1073/pnas.69.3.717

Chen, H., Sun, J., Li, S., Cui, Q., Zhang, H., Xin, F., et al. (2016). An ACC oxidase gene essential for cucumber carpel development. *Mol. Plant* 9, 1315–1327. doi: 10.1016/j.molp.2016.06.018

Galun, E. (1961). Study of the inheritance of sex expression in the cucumber. The interaction of major genes with modifying genetic and non-genetic factors. *Genetica* 32, 134–163. doi: 10.1007/BF01816091

Gao, P., Sheng, Y., Luan, F., Ma, H., and Liu, S. (2015). RNA-Seq transcriptome profiling reveals differentially expressed genes involved in sex expression in melon. *Crop Sci.* 55, 1686–1695. doi: 10.2135/cropsci2014.06.0444

Gu, H., Wang, D., Li, X., He, C., Xu, Z., and Bai, S. (2011). Characterization of an ethylene-inducible, calcium-dependent nuclease that is differentially expressed in cucumber flower development. *New Phytol.* 192, 590–600. doi: 10.1111/j.1469-8137.2011.03825.x

Hu, B., Li, D., Liu, X., Qi, J., Gao, D., Zhao, S., et al. (2017). Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 system. *Mol. Plant* 10, 1575–1578. doi: 10.1016/j.molp.2017.09.005

Ji, G., Zhang, J., Zhang, H., Sun, H., Gong, G., Shi, J., et al. (2016). Mutation in the gene encoding 1-aminocyclopropane-1-carboxylate synthase 4 *(CitACS4)* led to andromonoecy in watermelon. *J. Integr. Plant Biol.* 58, 762–765. doi: 10.1111/jipb.12466

Ji, G., Zhang, J., Gong, G., Shi, J., Zhang, H., Ren, Y., et al. (2015). Inheritance of sex forms in watermelon (*Citrullus lanatus*). *Sci. Hortic.* 193, 367–373. doi: 10.1016/j.scienta.2015.07.039

Kamachi, S., Sekimoto, H., Kondo, N., and Sakai, S. (1997). Cloning of a cDNA for a 1-aminocyclopropane-1-carboxylate Synthase that is expressed during development of female flowers at the apices of *Cucumis sativus* L. *Plant Cell Physiol.* 38, 1197–1206. doi: 10.1093/oxfordjournals.pcp.a029106

Kenigsbuch, D., and Cohen, Y. (1987). Inheritance of gynoecious sex type in muskmelon. *Cucurbit Genet. Coop. Rep.* 10, 47–48.

Kenigsbuch, D., and Cohen, Y. (1990). The inheritance of gynoecy in muskmelon. *Genome* 33, 317–320. doi: 10.1139/g90-049

Klee, H. J., and Giovannoni, J. J. (2011). Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* 45, 41–59. doi: 10.1146/annurev-genet-110410-132507

Knopf, R. R., and Trebitsh, T. (2006). The female-specific *CS-ACS1G* gene of cucumber. A case of gene duplication and recombination between the non-sex-specific *1-aminocyclopropane-1-carboxylate synthase* gene and a *branched-chain amino acid transaminase* gene. *Plant Cell Physiol.* 47, 1217–1228. doi: 10.1093/pcp/pcj092

Kubicki B. (1969d). Investigations on sex determination in cucumbers (*Cucumis sativus* L.). VI. Androecism. *Genet. Pol.* 10, 87–99.

Kubicki, B. (1969a). Investigations of sex determination in cucumber (*Cucumis sativus* L.), IV. Multiple alleles of locus *acr*. *Genet. Pol.* 10, 23–68.

Kubicki, B. (1969b). Investigations of sex determination in cucumber (*Cucumis sativus* L.), V. Genes controlling intensity of femaleness. *Genet. Pol.* 10, 69-86.

Kubicki, B. (1969c). Investigations of sex determination in cucumber (*Cucumis sativus* L.), VII. Trimonoecism. *Genet. Pol.* 10, 123–143.

Kubicki, B. (1974). New sex types in cucumber and their uses in breeding work. *Proc. XIX Intl. Hort. Congr.* 3, 475–485.

Kubicki, B. (1980). Investigation on sex determination in cucumber *Cucumis saticus* L. IX. Induced mutant with the recessive character of gynoecism. *Genet. Pol.* 21, 409–423.

Lai, Y., Shen, D., Zhang, W., Zhang, X., Qiu, Y., Wang, H., et al. (2018a). Temperature and photoperiod changes affect cucumber sex expression by different epigenetic regulations. *BMC Plant Biol.* 18, 268. doi: 10.1186/s12870-018-1490-3

Lai, Y., Zhang, W., Zhang, X., Shen, D., Wang, H., Qiu, Y., et al. (2018b). Integrative analysis of transcriptomic and methylomic data in photoperiod-dependent regulation of cucumber sex expression. *G3.* 8, 3981–3991. doi: 10.1534/g3.118.200755

Lai, Y., Zhang, X., Zhang, W., Shen, D., Wang, H., Xia, Y., et al. (2017). The association of changes in DNA methylation with temperature-dependent sex determination in cucumber. *J. Exp. Bot.* 68, 2899–2912. doi: 10.1093/jxb/erx144

Latrasse, D., Rodriguez-Granados, N. Y., Velucham, A., Mariappan, K. G., Bevilacqua, C., Crapart, N., et al. (2017). The quest for epigenetic regulation underlying unisexual flower development in *Cucumis melo*. *Epigenetics Chromatin* 10, 22. doi: 10.1186/s13072-017-0132-6

Li, Z., Huang, S., Liu, S., Pan, J., Zhang, Z., Tao, Q., et al. (2009). Molecular isolation of the *M* gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants. *Genetics* 182, 1381–1385. doi: 10.1534/genetics.109.104737

Li, Z., Wang, S., Tao, Q., Pan, J., Si, L., Gong, Z., et al. (2012). A putative positive feedback regulation mechanism in *CsACS2* expression suggests a modified model for sex determination in cucumber (*Cucumis sativus* L.). *J. Exp. Bot.* 63, 4475–4484. doi: 10.1093/jxb/ers123

Little, H. A., Papadopoulou, E., Hammar, S. A., and Grumet, R. (2007). The influence of ethylene perception on sex expression in melon (*Cucumis melo* L.) as assessed by expression of the mutant ethylene receptor, *At-etr1-1*, under the control of constitutive and floral targeted promoters. *Sex Plant Reprod.* 20, 123–136. doi: 10.1007/s00497-007-0049-5

Liu, M., Pirrello, J., Chervin, C., Roustan, J-P., and Bouzayen, M. (2015). Ethylene control of fruit ripening: revisiting the complex network of transcriptional regulation. *Plant Physiol.* 169, 2380–2390. doi: 10.1104/pp.15.01361

Ma, W-J., and Pannell, J. R. (2016). Sex determination: separate sexes are a double turnoff in melons. *Curr. Biol.* 26, R157–R179. doi: 10.1016/j.cub.2015.12.026

Manzano, S., Aguado, E., Martínez, C., Megías, Z., García, A., and Jamilena, M. (2016). The ethylene biosynthesis gene *CitACS4* regulates monoecy/andromonoecy in watermelon (*Citrullus lanatus*). *PLoS One* 11, e0154362. doi: 10.1371/journal.pone.0154362

Manzano, S., Martínez, C., Domínguez, V., Avalos, E., Garrido, D., Gómez, P, et al. (2010a). A major gene conferring reduced ethylene sensitivity and

maleness in *Cucurbita pepo*. *J. Plant Growth Regul.* 29, 73–80. doi: 10.1007/s00344-009-9116-5

Manzano, S., Martínez, C., García, J. M., Megías, Z., and Jamilena, M. (2014). Involvement of ethylene in sex expression and female flower development in watermelon (*Citrullus lanatus*). *Plant Physiol. Biotech.* 85, 96–104. doi: 10.1016/j.plaphy.2014.11.004

Manzano, S., Martínez, C., Gómez, P., Garrido, D., and Jamilena, M. (2010b). Cloning and characterisation of two CTR1-like genes in *Cucurbita pepo*: regulation of their expression during male and female flower development. *Sex Plant Reprod.* 23, 301–313. doi: 10.1007/s00497-010-0140-1

Manzano, S., Martínez, C., Megías, Z., Garrido, D., and Jamilena, M. (2013). Involvement of ethylene biosynthesis and signalling in the transition from male to female flowering in the monoecious *Cucurbita pepo*. *J. Plant Growth Regul.* 32, 789–798. doi: 10.1007/s00344-013-9344-6

Manzano, S., Martínez, C., Megías, Z., Gómez, P., Garrido, D., and Jamilena, M. (2011). The role of ethylene and brassinosteroids in the control of sex expression and flower development in *Cucurbita pepo*. *Plant Growth Regul.* 65, 213–221. doi: 10.1007/s10725-011-9589-7

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transponson-induced epigenetic change leads to sex determination in melon. *Nature*, 461, 1135–1138. doi: 10.1038/nature08498

Martínez, C., Manzano, S., Megías, Z., Barrera, A., Boualem, A., Garrido, D., et al. (2014). Molecular and functional characterization of *CpACS27A* gene reveals its involvement in monoecy instability and other associated traits in squash (*Cucurbita pepo* L.). *Planta* 239, 1201–1215. doi: 10.1007/s00425-014-2043-0

Miao, M., Yang, X., Han, X., and Wang, K. (2011). Sugar signalling is involved in the sex expression response of monoecious cucumber to low temperature. *J. Exp. Bot.* 62, 797–804. doi: 10.1093/jxb/erq315

Mibus, H., and Tatlioglu, T. (2004). Molecular characterization and isolation of the *F/f* gene for femaleness in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 109, 1669–1676. doi: 10.1007/s00122-004-1793-7

Niu, H., Liu, X., Tong, C., Wang, H., Li S., Lu, L., et al. (2018). The *WUSCHEL-related homeobox1* gene of cucumber regulates reproductive organ development. *J. Exp. Bot.* 69, 5373–5387. doi: 10.1093/jxb/ery329

Owens, K. W., Peterson, C. E., and Tolla, G. E. (1980). Production of hermaphrodite flowers on gynoecious muskmelon by silver nitrate and aminoethoxyvinylglycine. *HortScience* 15, 654–655.

Pan, J., Wang, G., Wen, H., Du, H., Lian, H., He, H., et al. (2018). Differential gene expression caused by the *F* and *M* loci provides insight into ethylene-mediated female flower differentiation in cucumber. *Front. Plant Sci.* 9, 1091. doi: 10.3389/fpls.2018.01091

Pannell, J. R. (2017). Plant sex determination. *Curr. Biol.* 27, R191–R197. doi: 10.1016/j.cub.2017.01.052

Papadopoulou, E., and Grumet, R. (2005). Brassinosteroid-induced femaleness in cucumber and relationship to ethylene production. *HortScience* 40, 1763–1767. doi: 10.1109/ISPDC.2005.37

Pawełkowicz, M. E., Skarzyńska, A., Pląder, W., and Przybecki, Z. (2019a). Genetic and molecular bases of cucumber (*Cucumis sativus* L.) sex determination. *Mol. Breeding* 39, 50. doi: 10.1007/s11032-019-0959-6

Pawełkowicz, M., Osipowski, P., Wojcieszek, M. Wóycicki, R., Witkowicz, J., Hincha, D., et al. (2012). Identification and characterization of genes connected with flower morphogenesis in cucumber. *BioTechnologia* 93, 123–134. doi: 10.5114/bta.2012.46577

Pawełkowicz, M., Pryszcz, L., Skarzyńska, A., Wóycicki, R. K., Posyniak, K., Rymuszka, J., et al. (2019b). Comparative transcriptome analysis reveals new molecular pathways for cucumber genes related to sex determination. *Plant Reprod.* 32, 193–216. doi: 10.1007/s00497-019-00362-z

Peterson, C. E., and Andher, E. D. (1960). Induction of staminate flowers on gynoecious cucumbers with gibberellin A3. *Science* 131, 1673–1676. doi: 10.1126/science.131.3414.1673

Poole, C. F., and Grimball, P. C. (1939). Inheritance of new sex forms in *Cucumis melo* L. *J. Hered.* 30, 21–25. doi: 10.1093/oxfordjournals.jhered.a104626

Robinson, R. W., Munger, H. M., Whitaker, T. W., and Bohn, G. W. (1976). Genes of the Cucurbitaceae. *HortScience* 11, 554–568.

Rosa, J.T. (1928). The inheritance of flower types in *Cucumis* and *Citrullus*. *Hilgardia* 3, 233–250. doi: 10.3733/hilg.v03n09p233

Rudich, J., Halevy, A. H., and Kedar, N. (1969). Increase in femaleness of three cucurbits by treatment with Ethrel, an ethylene releasing compound. *Planta* 86, 69–76. doi: 10.1007/BF00385305

Rudich, J., Halevy, A. H., and Kedar, N. (1972). Ethylene evolution from cucumber plants as related to sex expression. *Plant Physiol.* 49, 998–999. doi: 10.1104/pp.49.6.998

Saito, S., Fujii, N., Miyazawa, Y., Yamasaki, S., Matsuura, S., Mizusawa, H., et al. (2007). Correlation between development of female flower buds and expression of the *CS-ACS2* gene in cucumber plants. *J. Exp. Bot.* 58, 2897–2907. doi: 10.1093/jxb/erm141

Shifriss, O. (1961). Sex control on cucumbers. *J. Hered.* 52, 5–12. doi: 10.1093/oxfordjournals.jhered.a107021

Song, J., Zhang, Y., Song, S., Su, W., Chen, R., Sun, G., et al. (2018). Comparative RNA-Seq analysis on the regulation of cucumber sex differentiation under different ratios of blue and red light. *Bot. Stud.* 59, 21. doi: 10.1186/s40529-018-0237-7

Switzenberg, J. A., Little, H. A., Hammar, S. A., and Grumet, R. (2014). Floral primordia-targeted *ACS* (*1-aminocyclopropane-1-carboxylate synthase*) expression in transgenic *Cucumis melo* implicates fine tuning of ethylene production mediating unisexual flower development. *Planta* 240, 797–808. doi: 10.1007/s00425-014-2118-y

Takahashi, H., and Jaffe, M. J. (1984). Thigmomorphogenesis: the relationship of mechanical perturbation to elicitor-like activity and ethylene production. *Physiol. Plant.* 61, 405–411. doi: 10.1111/j.1399-3054.1984.tb06347.x

Takahashi, H., Saito, T., and Suge, H. (1982). Intergeneric translocation of floral stimulus across a graft in monoecious Cucurbitaceae with special reference to the sex expression of flowers. *Plant Cell Physiol.* 23, 1–9. doi: 10.1094/Phyto-72-1598

Tan, J., Tao, Q., Niu, H., Zhang, Z., Li, D., Gong, Z., et al. (2015). A novel allele of *monoecious* (*m*) locus is responsible for elongated fruit shape and perfect flowers in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 128, 2483–2493. doi: 10.1007/s00122-015-2603-0

Tanurdzic, M., and Banks, J. A. (2004). Sex-determining mechanisms in land plants. *Plant Cell* 16, S61–S71. doi: 10.1105/tpc.016667

Tao, Q., Niu, H., Wang, Z., Zhang, W., Wang, H., Wang, S., et al. (2018). Ethylene responsive factor ERF110 mediates ethylene-regulated transcription of a sex determination-related orthologous gene in two *Cucumis* species. *J. Exp. Bot.* 69, 2953–2965. doi:10.1093/jxb/ery128. doi: 10.1093/jxb/ery128

Trebitsh, T., Rudich, J., and Riov, J. (1987). Auxin, biosynthesis of ethylene and sex expression in cucumber (*Cucumis sativus*). *Plant Growth Regul.* 5, 105–113. doi: 10.1007/BF00024738

Trebitsh, T., Staub, J. E., and O'Neill, S. D. (1997). Identification of a *1-aminocyclopropane-1-carboxylic acid synthase* gene linked to the *female* (*F*) locus that enhances female sex expression in cucumber. *Plant Physiol.* 113, 987–995. doi: 10.1104/pp.113.3.987

Tsao, T. H. (1988). Sex expression in flowering. *Acta Phyto. Sin.* 14, 203–207.

Wang, C., Xin, M., Zhou, X., Liu, W., Liu, D., and Qin, Z. (2018). Transcriptome profiling reveals candidate genes associated with sex differentiation induced by night temperature in cucumber. *Sci. Hort.* 232, 162–169. doi: 10.1016/j.scienta.2017.12.018

Wang, D., Li, F., Duan, Q., Han, T., Xu, Z., and Bai, S. (2010). Ethylene perception is involved in female cucumber flower development. *Plant J.* 61, 862–872. doi: 10.1111/j.1365-313X.2009.04114.x

Wang, H., Yu, C., Fan, P., Bao, B., Li, T., and Zhu, Z. (2014). Identification of two cucumber putative silicon transporter genes in *Cucumis sativu*s. *J. Plant Growth Regul.* 34, 332–338. doi: 10.1007/s00344-014-9466-5

Wang, R., Lin, Y., Jin, Q., Yao, C., Zhong, Y., and Wu, T. (2019). RNA-Seq analysis of gynoecious and weak female cucumber revealing the cell cycle pathway may regulate sex determination in cucumber. *Gene* 687, 289–297. doi: 10.1016/j.gene.2018.11.071

Win, K. T., Zhang, C., Silva, R. R., Lee, J. H., Kim, Y-C., and Lee, S. (2019). Identification of quantitative trait loci governing subgynoecy in cucumber. *Theor. Appl. Genet.* 132, 1505–1521. doi: 10.1007/s00122-019-03295-3

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

Yamasaki, S., Fujii, N., Matsuura, S., Mizusawa, H., and Takahashi, H. (2001). The *M* locus and ethylene-controlled sex determination in andromonoecious cucumber plants. *Plant Cell Physiol.* 42, 608–619. doi: 10.1093/pcp/pce076

Yang, S. F., and Hoffman, N. E. (1984). Ethylene biosynthesis and its regulation in higher plants. *Annu. Rev. Plant Physiol.* 35, 155–189. doi: 10.1146/annurev. pp.35.060184.001103

Yin, T., and Quinn, J. A. (1995). Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am. J. Bot.* 82, 1537–1546. doi: 10.2307/2446182

Zhang, J., Shi, J., Ji, G., Zhang, H., Gong, G., Guo, S., et al. (2017a). Modulation of sex expression in four forms of watermelon by gibberellin, ethephone and silver nitrate. *Hort. Plant J.* 3, 91–100. doi: 10.1016/j.hpj.2017.07.010

Zhang, J., Guo, S., Ji, G., Zhao, H., Sun, H., Ren, Y., et al. (2019). A unique chromosome translocation disrupting *ClWIP1* leads to gynoecy in watermelon. *Plant J.* doi: 10.1111/tpj.14537

Zhang, Y., Liu, B., Yang, S., An, J., Chen, C., Zhang, X., et al. (2014a). 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

Zhang, Y., Zhang, X., Liu, B., Wang, W., Liu, X., Chen, C., et al. (2014b). A GAMYB homologue CsGAMYB1 regulates sex expression of cucumber via an ethylene independent pathway. *J. Exp. Bot.* 65, 3201–3213. doi: 10.1093/jxb/eru176

Zhang, Y., Zhao, G., Li, Y., Mo, N., Zhang, J., and Liang, Y. (2017b). Transcriptomic analysis implies that GA regulates sex expression via ethylene-dependent and ethylene-independent pathways in cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* 8, 10. doi: 10.3389/fpls.2017.00010

Zhang, Z., Mao, L., Chen, H., Bu, F., Li, G., Sun, J., et al. (2015). Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell* 27, 1595–1604. doi: 10.1105/tpc.114.135848

Zhang, Z., Zhang, H., Quan, R., Wang, X., and Huang, R. (2009). Transcriptional regulation of the ethylene response factor LeERF2 in the expression of ethylene biosynthesis genes controls ethylene production in tomato and tobacco. *Plant Physiol.* 150, 365–377. doi: 10.1104/pp.109.135830

Zhou, Y., Ahammed, G. J., Wang, Q., Wu, C., Wan, C., and Yang, Y. (2018). Transcriptomic insights into the blue light-induced female floral sex expression in cucumber (*Cucumis sativus* L.). *Sci. Rep.* 8, 14261. doi: 10.1038/s41598-018-32632-7

A Single Nucleotide Deletion in an ABC Transporter Gene Leads to a Dwarf Phenotype in Watermelon

Huayu Zhu†, Minjuan Zhang†, Shouru Sun, Sen Yang, Jingxue Li, Hui Li, Huihui Yang, Kaige Zhang, Jianbin Hu, Dongming Liu and Luming Yang*

College of Horticulture, Henan Agricultural University, Zhengzhou, China

Dwarf habit is one of the most important traits in crop plant architecture, as it can increase plant density and improved land utilization, especially for protected cultivation, as well as increasing lodging resistance and economic yield. At least four dwarf genes have been identified in watermelon, but none of them has been cloned. In the current study, the *Cldw-1* gene was primary-mapped onto watermelon chromosome 9 by next-generation sequencing-aided bulked-segregant analysis (BSA-seq) of F_2 plants derived from a cross between a normal-height line, WT4, and a dwarf line, WM102, in watermelon. The candidate region identified by BSA-seq was subsequently validated and confirmed by linkage analysis using 30 simple sequence repeat (SSR) markers in an F_2 population of 124 plants. The *Cldw-1* gene was further fine-mapped by chromosome walking in a large F_2 population of 1,053 plants and was delimited into a candidate region of 107.00 kb. Six genes were predicted to be in the candidate region, and only one gene, *Cla010337*, was identified to have two single nucleotide polymorphisms (SNPs) and a single nucleotide deletion in the exons in the dwarf line, WM102. A derived cleaved amplified polymorphic sequence (dCAPS) marker was developed from the single nucleotide deletion, co-segregated with the dwarf trait in both the F_2 population and a germplasm collection of 165 accessions. *Cla010337* encoded an ATP-binding cassette transporter (ABC transporter) protein, and the expression levels of *Cla010337* were significantly reduced in all the tissues tested in the dwarf line, WM102. The results of this study will be useful in achieving a better understanding of the molecular mechanism of the dwarf plant trait in watermelon and for the development of marker-assisted selection (MAS) for new dwarf cultivars.

Keywords: watermelon, dwarf, BSA-seq, SSR marker, fine mapping, ABC transporter

INTRODUCTION

Cultivated watermelon (*Citrullus lanatus* var. *lanatus* (Thunb.) Matsum. & Nakai) is an important horticultural crop in the Cucurbitaceae family and is one of the most widely consumed fresh fruits worldwide. The diploid genome of watermelon is relatively small (~425 Mb; Guo et al., 2012), consisting of twenty-two chromosomes (2n = 2x = 22). The draft genome of the East Asian watermelon cultivar 97103 was sequenced and assembled in 2012 using next-generation sequencing (NGS) technology (Guo et al., 2012). A large number of SSR and insertion/deletion

(Indel) markers have been developed from the assembly of the watermelon genome (Ren et al., 2012; Zhu et al., 2016). The rapid development of genetic and genomic resources in watermelon was greatly facilitated by gene and QTL mapping. A number of genes controlling important traits in watermelon have been mapped and cloned, such as fruit shape (Dou et al., 2018), flesh color (Zhang et al., 2017), the tonoplast sugar transporter gene (*ClTST2*) (Ren et al., 2018), and lobed leaf shape (Wei et al., 2017), as well as QTLs associated with resistance to different pathogens (Lambel et al., 2014; Kim et al., 2015; Ren et al., 2015; Branham et al., 2017).

Dwarf plant habit is an important trait in plant breeding, as reduced height is associated with lodging tolerance, increased economic sink size, and stable yield increases in many crops. A number of genes controlling plant height have been identified and mapped in the Cucurbitaceae family. To date, at least five genes have been identified for compact growth habit or dwarfism-related plant architecture in cucumber, including *compact* (*cp*) (Kauffman and Lower, 1976), *cp-2* (Kubicki et al., 1986), *supercompact* (*scp*) (Niemirowicz-Szczytt et al., 1996), *scp-1* (Wang et al., 2017), and *scp-2* (Hou et al., 2017). The *cp* mutant was fine-mapped to a 220 kb region at the end of chromosome 4 in cucumber (Li et al., 2011). The mutants *scp-1* and *scp-2* were identified as carrying mutations in the plant cytochrome P450 monooxygenase gene *CsCYP85A1* and the *de-etiolated-2* gene (*CsDET2*), respectively, which both play important roles in the brassinosteroid (BR) biosynthesis pathway (Hou et al., 2017; Wang et al., 2017). In melon, four recessive dwarfing genes have been reported, *si-1*, *si-2*, *si-3*, and *mdw1,* which were identified from four independent melon lines (Paris et al., 1984; Knavel, 1990; Hwang et al., 2014). Among them, the *si-1* plants exhibit a bush phenotype with an extremely compact growth habit and very short internode lengths (Denna, 1962), and the *si-1* gene is linked to the gene *yv* (*yellow virescent*) (Pitrat, 1991). The internode lengths in *si-2* and *si-3* plants are shorter but the plants are less compact than the *si-1* plants. In *si-2* plants, only the first internodes are short, with no effect on the length of the upper, later-formed internodes, whereas the internode lengths of *si-3* plants are reduced at all plant developmental stages. Of these four melon dwarfing genes, only *mdw1* has been mapped, onto a location between two gene markers in a 1.8 cM region on melon chromosome 7 (Hwang et al., 2014). In pumpkin, a major QTL, *qCmB2*, explained 21.39% of the phenotypic variation for a dwarf bush type and was mapped to a 0.42 Mb region using a high-density genetic map. The *Gibberellin* (*GA*) *20-oxidase* gene was identified as the possible candidate gene controlling vine growth (Zhang et al., 2015). Several dwarfing genes were also identified in watermelon, including two alleles, *dw-1* and *dw-1ˢ*, and two independent loci, *dw-2* and *dw-3* (Liu and Loy, 1972; Mohr and Sandhu, 1975; Huang et al., 1998). Recently, another dwarf gene, *dsh*, was identified in watermelon from a natural mutation that exhibited both small fruits and short internodes. The *dsh* gene was mapped onto a 27,800 kb long region on watermelon chromosome 7 by BSA-seq analysis and a *gibberellin 20-oxidase-like* gene was predicted as a possible candidate gene (Dong et al., 2018). Although a number of dwarf genes have been reported and mapped in the Cucurbitaceae family, only a few of them have been cloned, and little is known about the underlying molecular mechanisms.

Most of the dwarfing mutations in different crops have been identified as being associated with the loss of function of genes associated with biosynthesis or of response pathways of plant hormones that regulate cell elongation and division (Sasaki et al., 2002; Multani et al., 2003; Nomura et al., 2004; Pearce et al., 2011; Tamiru et al., 2015). Gibberellins (GAs) are well known for playing important roles in controlling plant height, and most GA-deficient or -insensitive mutants are characterized by reduced height and associated developmental phenotypic changes (Sakamoto et al., 2004). Another group of plant hormones, the BRs, have been widely identified as being involved in the regulation of plant height (Hong et al., 2004), and a number of BR-deficient or insensitive dwarf mutants have been isolated in several crops (Tamiru et al., 2015; Hou et al., 2017; Wang et al., 2017). In addition, Multani et al. (2003) identified a different mechanism by which plant height is controlled in dwarf mutants of maize and sorghum and found that the *br2* gene in a maize mutant and the *dw3* gene in a sorghum mutant encoded a protein responsible for the transport of the plant hormones auxins (Multani et al., 2003), indicating that auxins also play a role in the regulation of plant height. Furthermore, more recently identified plant hormones, the strigolactones (SLs), have been revealed as being associated with plant height regulation, controlling stem elongation independently of GAs (de Saint Germain et al., 2013). These findings suggest that the dwarfism trait in different crops may be controlled by various genes that operate *via* different molecular mechanisms and that the classification of dwarfing genes that regulate plant height in combination with desirable pleiotropic effects appropriate for particular crops will enable the breeding of better crop varieties in the near future.

The ATP-binding cassette (ABC) transporter gene family is one of the largest and most ubiquitous gene superfamilies, being present in plants and animals. ABC transporters can transport a wide range of molecules across various membrane types (Do et al., 2018). They are essential for plant development and have been identified as being involved in processes as diverse as gametogenesis, seed development, seed germination, organ formation, and secondary growth (Hwang et al., 2016). The functional ABC transporters are usually composed of two transmembrane domains (TMD) and two nucleotide-binding domains (NBD). Based on the conservation of the NBD sequence, the ABC transporter family can be divided into eight subfamilies (A-H). Of these, several members of the B subfamily (the ABC-B/multidrug resistance/P-glycoprotein, and ABCB/MDR/PGP subfamily), namely ABCB1, ABCB4, ABCB19, ABCB14, and ABCB15, have been identified as auxin transporters involved in plant height regulation (Noh et al., 2001; Kaneda et al., 2011; Kubes et al., 2012). However, the molecular mechanism by which ABCB genes regulate plant height is still unclear, as is whether there is crosstalk between auxins and other hormone pathways in plant height regulation.

In the current study, a dwarf watermelon line, WM102, carrying the single recessive gene, *Cldw-1*, was used to cross with a normal-height watermelon line, WT4. A BSA-seq strategy was used for primary mapping of *Cldw-1* onto the end of watermelon chromosome 9. Based on the whole-genome re-sequencing of the two parental lines, Indel and dCAPS markers were developed from the candidate region and were then used to genotype a

large F_2 mapping population. A mutant ABC transporter gene, *Cla010337*, shown to carry a single nucleotide deletion frameshift mutation in the coding region in the dwarf line, WM102, was identified as the candidate gene for *Cldw-1*.

MATERIALS AND METHODS

Plant Materials and Mapping Populations

WM102 is a dwarf inbred line, which was selected from 'Bush Sugar Baby' [accession code: Grif15898; provided by USDA-ARS Germplasm Resources Information Network (GRIN) (https://www.ars-grin.gov/Pages/Collections)] and had been manually self-pollinated for at least four generations. It exhibited short internodes in both the primary and secondary stems. To confirm the inheritance mode and to achieve fine mapping of the *dwarf1* (*Cldw-1*), WM102 was crossed with a normal-height inbred line, WT4, to generate a large F_2 mapping population. Phenotype scoring for plant height was carried out at 60 days of age, planting in a greenhouse in 2017 and 2018.

To investigate the allelic diversity of the *Cldw-1* gene in natural populations, the *Cldw-1* co-segregating markers dCAPS3 and Indel1 were used to screen a panel of 165 watermelon germplasms of worldwide collection. The plant heights of five plants of each of the 165 watermelon accessions were recorded as normal-height or dwarf plant at 30 and 60 days in the field in 2017 and 2018. Detail information on these accessions is provided in **Table S1**. All the watermelon germplasm and F_2 population were grown in the greenhouse at the Maozhuang Research Station of Henan Agricultural University (Zhengzhou, China).

Cytological Characterization of the Dwarf Trait

To compare the cytological characteristics of the internodes between the two parental lines, in autumn 2018, the eighth internodes of three different 45-day-old plants of WM102 and WT4 were separately fixed, washed, dehydrated, dewaxed, embedded, sectioned, and finally viewed using an Olympus-BX53 light microscope, as described previously (Yang et al., 2019). For the measurement of the cell size, three different embedded samples were sectioned and used for investigation. On average, a series of 10 to 20 paraffin sections (10-μm thick) was counted and compared.

In addition to paraffin sectioning, epoxy resin semi-thin sectioning was performed to observe the cell in the eighth internodes between the two parental lines, WM102 and WT4, in autumn 2018. Samples were fixed in 2.5% glutaraldehyde for 4 h at room temperature and rinsed three times with phosphate buffer, post-fixed with 1% osmic acid for 1.5 h at room temperature, washed three times in distilled water, dehydrated in an acetone series, and finally, embedded in LR White acrylic resin. Semi-thin sections were taken and supported on Formvar-coated gold grids and then observed with the Olympus-BX53 light microscope.

Bulk Pooling and BSA-Seq Analysis

For next-generation sequencing-aided bulked-segregant analysis (BSA-seq), 20 dwarf plants and 20 normal-height plants from the F_2 mapping populations in the greenhouse were selected at random

for DNA extraction. Unexpanded young leaves from these plants were collected into 1.5 mL microcentrifuge tubes, lyophilized in a freeze dryer, and ground into fine powder. Genomic DNA was extracted using the CTAB method (Murray and Thompson, 1980).

The dwarf plant bulk (D-bulk) and the normal-height plant bulk (N-bulk) were generated by pooling equal amounts of DNA from each of the dwarf plants and normal-height plants, respectively. A 5-ug DNA sample was then taken from each of the two bulks and from each of the two parental lines. These samples were used to construct paired-end sequencing libraries, which were sequenced on an Illumina HiSeq™ 2500 platform.

After removing adapter sequences, short reads, and low-quality reads, the clean reads from the two bulks were further rechecked for quality and then used for mapping to the watermelon reference genome 97103 (ftp://cucurbitgenomics.org/pub/cucurbit/genome/watermelon/97103/v1) with a Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). GATK (Genome Analysis Toolkit) was used to detect SNPs and small Indels between the two parental lines and the two bulks, respectively (McKenna et al., 2010). SNPs and small Indels from the alignments were called using Samtools, and the output was given in pileup format (Li et al., 2009). The quality score of the SNPs was assigned by Samtools to evaluate the reliability of SNP calling based on the Phred-scaled probability that the consensus is identical to the reference. The reliable SNPs and small Indels were noted and predicted using SnpEff software (Chen et al., 2009), and only the high-quality SNPs, with a minimum of five-sequence read depth were used for BSA-seq analysis.

Two methods were used to detect the genomic regions associated with dwarfism: the Euclidean Distance (ED) algorithm and SNP-Index analysis. The detailed calculation for ED was as described previously (Hil et al., 2013). The ED values were raised to a power of 5 ($ED5$) to decrease the noise generated by small variations in the estimations. A high ED value suggested that the SNPs in the genomic regions were closely associated with the target genes. The SNP-index is an association analysis method for finding significant differences in genotype frequency between the pools as indicated by ΔSNP-index; the same process as has been detailed previously was followed (Abe et al., 2012; Takagi et al., 2013). The SNP index is calculated as follows: SNP-index (Dwarf) = $\rho x/(\rho X + \rho x)$, SNP-index (WT) = $\rho x/(\rho X + \rho x)$, ΔSNP-index = SNP-index (Dwarf) - SNP-index (WT). Dwarf and WT represent the dwarf plant bulk and the normal-height plant bulk of the filial generation, respectively. ρX and ρx indicate the number of reads of the alleles in the WT and the dwarf parent lines appearing in their pools, respectively. The difference in each locus between the two pools can be observed through the ΔSNP-index. With respect to qualitative characterization, the correlation threshold is the theoretical ΔSNP-index value of the corresponding population, and the correlation threshold of the F_2 population is 0.67. The regions over the threshold were considered as the associated candidate regions.

Marker Development and Linkage Analysis

The SSR markers in the candidate region, which were developed from a genome-wide SSR identification (Zhu et al., 2016), were used to screen for polymorphism between the two parental lines.

The good polymorphic markers were then used for use in linkage analysis of an F_2 mapping population containing 124 plants to validate the BSA-seq result. After the initial localization of *Cldw-1* on chromosome 9, a scaffold-based chromosome-walking strategy was undertaken to identify markers that were more closely linked. SNPs and Indels were explored based on the sequence difference between WM102 and WT4 by comparing the genome re-sequencing data. For Indels, only those polymorphisms with ≥3 bp differences were selected for primer design with Primer3 software (http://primer3.ut.ee/). For SNP genotyping, dCAPS markers were developed with dCAPS Finder 2.0 (Neff et al., 1998).

The PCR amplification of molecular markers and subsequent gel electrophoresis were conducted as described by Zhu et al. (2016). Linkage analysis of the *Cldw-1* locus with molecular markers was performed with the Kosambi mapping function in JoinMap 4.0, using the threshold logarithm$_{10}$ of the odds (LOD) score of 5.0.

RNA Extraction, CDNA Synthesis, and Qrt-PCR Analysis

Total RNA was isolated from different tissues using the Plant RNA kit (Omega, USA) according to the manufacturer's protocol. The quality and quantity of RNA samples were assessed on a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). cDNA was synthesized using the SuperScript III Reverse Transcriptase (Invitrogen, USA) after the elimination of genomic DNA from the RNA. The expression pattern of the candidate gene was examined by performing qRT-PCR analysis of several tissues from WT4 and WM102, including the hypocotyl and the entire root system, the first leaf near the apical meristems, and the entire stems of 35-day-old plants. All experiments were performed with three biological and four technical replicates. The *ClCAC* (*Clathrin adaptor complex subunit*, Cla020794) gene, which has been validated with respect to stable expression in different watermelon organs and tissues under various conditions (Kong et al., 2014), was used as an internal control. qRT-PCR was performed using the SYBR Green PCR Master Mix (Applied Biosystems Inc., USA) in an iCycleriQTM 5 Multicolor Real-Time PCR detection system (Bio-Rad, USA). The threshold cycle (Ct) value of the *CAC* gene was subtracted from that of *Cldw-1* to obtain a ΔCt value. The Ct value of the control sample, WM102, was subtracted from the ΔCt value to obtain a ΔΔCt value. The fold changes in expression level relative to WM102 were expressed as $2^{-\Delta\Delta Ct}$. The *ClDW1*-specific qRTPCR primers were obtained using Primer3web (http://primer3.ut.ee/). We analyzed the primer specificity by comparing it with the watermelon genomic sequence. Gene-specific primers of *Cldw-1* and *ClCAC* are provided in **Table S2**. *Cldw-1* gene-specific qRT-PCR primers were developed, locating on the 11th exons, with a product size of 83 bp. The thermal cycler settings for qRT-PCR consisted of an initial hold at 95°C for 2 mins, 40 cycles of 95°C for 15 s and 60°C for 30 s, then 95°C for 15s, 60°C for 1 min, and 95°C for 15s.

Phylogenetic Analysis of CIDW-1

To investigate the phylogenetic relationships of the watermelon ClDW-1 protein with members of the identified subfamily B of the ABC transporter superfamily, all 20 members of the ABCB subfamily in the *Arabidopsis* genome (https://www.arabidopsis.

org/), with another two ABCB proteins, ZmBR2 (GenBank accession number: AY366085) and SbDW3 (GenBank accession number: AY372819) (Multani et al., 2003), were used in this study. The full-length protein sequences of all the subfamily B members were aligned using the CLUSTX program (Chenna et al., 2003), and the alignment report was visualized by Gendoc. The phylogenetic tree was constructed using the maximum likelihood method in MEGA 5 with a bootstrap of 1,000 replicates (Tamura et al., 2011).

Statistical Analysis

The χ^2-test for goodness-of-fit was used to test for deviation of the observed data from the theoretically expected segregation for the dwarf plants in the F_2 population from the WM102 × WT4 cross. Student's *t*-test was used for comparison of fold change expression levels between WM102 and WT4. A difference was considered to be statistically significant when $P < 0.05$. Summary statistics are presented as mean ± standard deviation (SD).

RESULTS

Morphological Characterization and Inheritance of Dwarf Trait in Watermelon

Comparative morphological characterization between the two parental lines was conducted by measuring the hypocotyl length, main stem length, and internode length and comparing their values at different developmental stages in spring and autumn 2018. The hypocotyl length of 15-day-old seedings in WM102 was about 1.52 cm, which was almost half of that in WT4 (**Table 1** and **Figure S1**). Though the plant height and internode length were varied at different developmental stages in various environments, the relative length of plant height and internode length were significantly reduced in the dwarf inbred line, WM102, being almost half that in the normal-height line, WT4 (**Table 1**, **Figure 1** and **Figure S1**). However, there were no significant differences between the two parental lines in terms of the length of the first three nodes, which were very short and clustered together in both accessions. The difference in internode length and plant height between the two parental lines became more evident from the fourth node (**Figures 1A, B**). In addition, the lateral internode and tendril lengths were also significantly shorter in the dwarf line WM102. To further compare the cell structure of the internodes of WM102 and WT4 plants, the cell sizes of the eighth internodes of 45-day-old plants were viewed under a light microscope by semi-thin sectioning and paraffin sectioning. The cytological characterization

TABLE 1 | Morphological characterization of two parental lines in spring 2018.

Trait	Stage	WT4	WM102
Hypocotyl length (cm)	15-day-old	3.42 ± 0.59	1.52 ± 0.48
Plant height (cm)	30-day-old	185.20 ± 10.52	86.80 ± 7.43
	60-day-old	235.30 ± 6.26	128.72 ± 8.1
Internode length (cm)	30-day-old	5.54 ± 0.11	3.27 ± 0.31
	60-day-old	6.36 ± 0.11	4.16 ± 0.22

FIGURE 1 | Morphological characterization of two parental lines, WM102 and WT4. **(A)** 35-day-old plants of WM102 and WT4 grown in the growth chamber. Bar = 5 cm. **(B)** From left to right, the first three internodes and then the 4th to 8th internodes of WM102 and WT4. In each group, the left represents inbred line WM102, and the right represents inbred line WT4. Bar = 1 cm. **(C)** 45-day-old plants of WM102 and WT4 grown in the field.

showed that the cells, especially the parenchyma cells, in the internodes of WM102 plants were significantly smaller than in those of WT4 (**Figure 2**), a finding consistent with the shorter internode.

The phenotype of plant height for the F_2 population was investigated in spring and autumn 2018 (named 2018S-F2 and 2018A-F2, respectively), and genetic analysis revealed that the segregation of normal-height and dwarf plants was consistent with the segregation of a single recessive gene in both 2018S-F2 and 2018A-F2, which was named as *Cldw-1* (**Table 2**).

PRELIMINARY MAPPING BY BSA-SEQ ANALYSIS

We selected 20 dwarf plants and 20 normal-height plants at random from the F_2 population for pooling into two bulks: the dwarf plant bulk (D-bulk) and the normal-height plant bulk (N-bulk). The DNA from each of the two bulks and each of the two parental lines was used for paired-end library construction and next-generation sequencing (NGS). After filtering the raw reads, the two re-sequencing parental lines generated 79,907,647

and 73,483,217 clean reads totaling 23.94 and 22.02 Gb for WT4 and WM102, respectively. In addition, a total of 33.13 Gb clean reads were obtained (15.39 Gb from the D-bulk and 17.74 Gb from the N-bulk) with an average 34 × depth of the draft watermelon reference genome. All the clean reads of the two parental lines and the two bulks were then mapped to watermelon reference genome 97103, with more than 80% of the total reads being precisely mapped onto specific watermelon chromosomes. In total, 401,140 SNPs were detected after removing SNPs with low coverage or discrepancy between the parental lines and the bulks, and only 115,867 high-quality SNPs were finally retained.

To identify the genomic region associated with the dwarf phenotype, we first used the ED algorithm to measure the allelic segregation of SNPs between the D-Bulk and the N-Bulk. There were two statistically significant peaks of ED detected on chromosome 9, which were located close together in a large region between 19,710,000 and 26,970,000 bp (**Figure 3A**). Then, the SNP-Index of each SNP locus was also calculated between the two bulks using the high-quality SNPs, but no significant region was identified to be associated with the dwarf phenotype. However, two obvious adjacent peaks were still detected
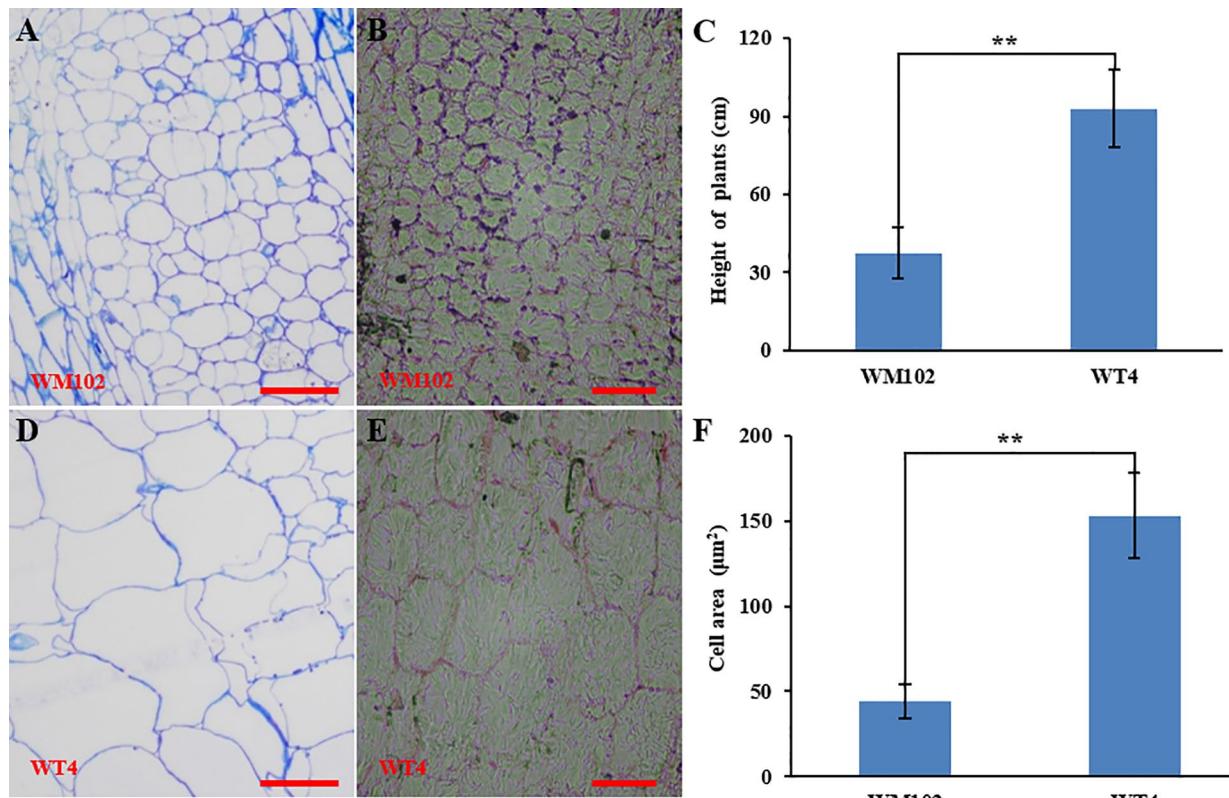
FIGURE 2 | Cytological characterization of the parental lines, WM102 and WT4. **(A**, **B)** Paraffin sections and semithin sections of the eighth internodes in 45-day-old plants of WM102. **(C)** Quantification of the plant height of 45-day-old plants of WM102 and WT4 in autumn 2018. Asterisks indicate that the WM102 plants are significantly shorter than WT4 plants (**P < 0.01). **(D**, **E)** Paraffin sections and semithin sections for the eighth internodes in 45-day-old plants of WT4. **(F)** Quantifications of cell size of the eighth internodes in three different WM102 and WT4 plants in autumn 2018. Asterisks indicate that the cell size in the WM102 plants is significantly smaller than that in WT4 (**P < 0.01). Scale bars: 100 μm.

TABLE 2 | Segregation analysis of the dwarf plant in the F_2 population.

Population	Total	Dwarf type	Normal type	Expected ratio	χ^2	P
F_1	15	0	15	–	–	–
2018S-F_2	538	140	398	3:1	0.2478	>0.05
2018A-F_2	515	144	371	3:1	2.2531	>0.05

under the significant threshold, and these were located in the same candidate regions as had been detected by ED analysis (**Figure 3B**). This indicated that the two candidate regions probably contained the candidate gene regulating plant height in watermelon. One peak was located between 22,540,000 and 22,810,000 bp, and the other peak was located from 28,640,000 to 31,590,000 bp on chromosome 9.

Validation of the BSA-Seq Result and Fine Mapping of the *Cldw-1* Gene

A total of 52 SSR markers in the candidate regions on chromosome 9 were selected from a genome-wide SSR marker identification in watermelon (Zhu et al., 2016) for screening polymorphisms, of which 22 showed clear bands and polymorphism between

WT4 and WM102. A further 18 SSR markers were selected from these for genotyping a small F_2 population of 124 plants, and the *Cldw-1* gene was mapped between ClSSR25954 and ClSSR26018 within a genetic distance of 8.5 cM, indicating that the BSA-seq mapping result was consistent with the linkage analysis. To further shorten the genetic distance of markers flanking the target gene, a chromosome-walking strategy was used, and additional SSR markers were selected from the candidate region for polymorphism screening. After three rounds of chromosome walking, 30 SSR markers were used for the final map construction, and the *Cldw-1* gene was still mapped between ClSSR25954 and ClSSR26018 within a genetic distance of 8.3cM (**Figure 4A**). Detail information on these SSR markers is provided in **Table S2**.

To check the primary mapping results, we selected an additional 414 F_2 plants, using eight SSR markers closely linked to *Cldw-1* for genotyping. We found that the two closely flanking markers of *Cldw-1* were slightly different in this larger population, with the *Cldw-1* gene being flanked by ClSSR26018 and ClSSR26045 with map distances from the target gene of 2.7 and 0.7 cM (**Figure 4B**), respectively. This discrepancy between the results from the two different mapping populations was probably because the primary mapping of *Cldw-1* used a small mapping population and a high density of markers, which led to the marker order not precisely reflecting their genetic
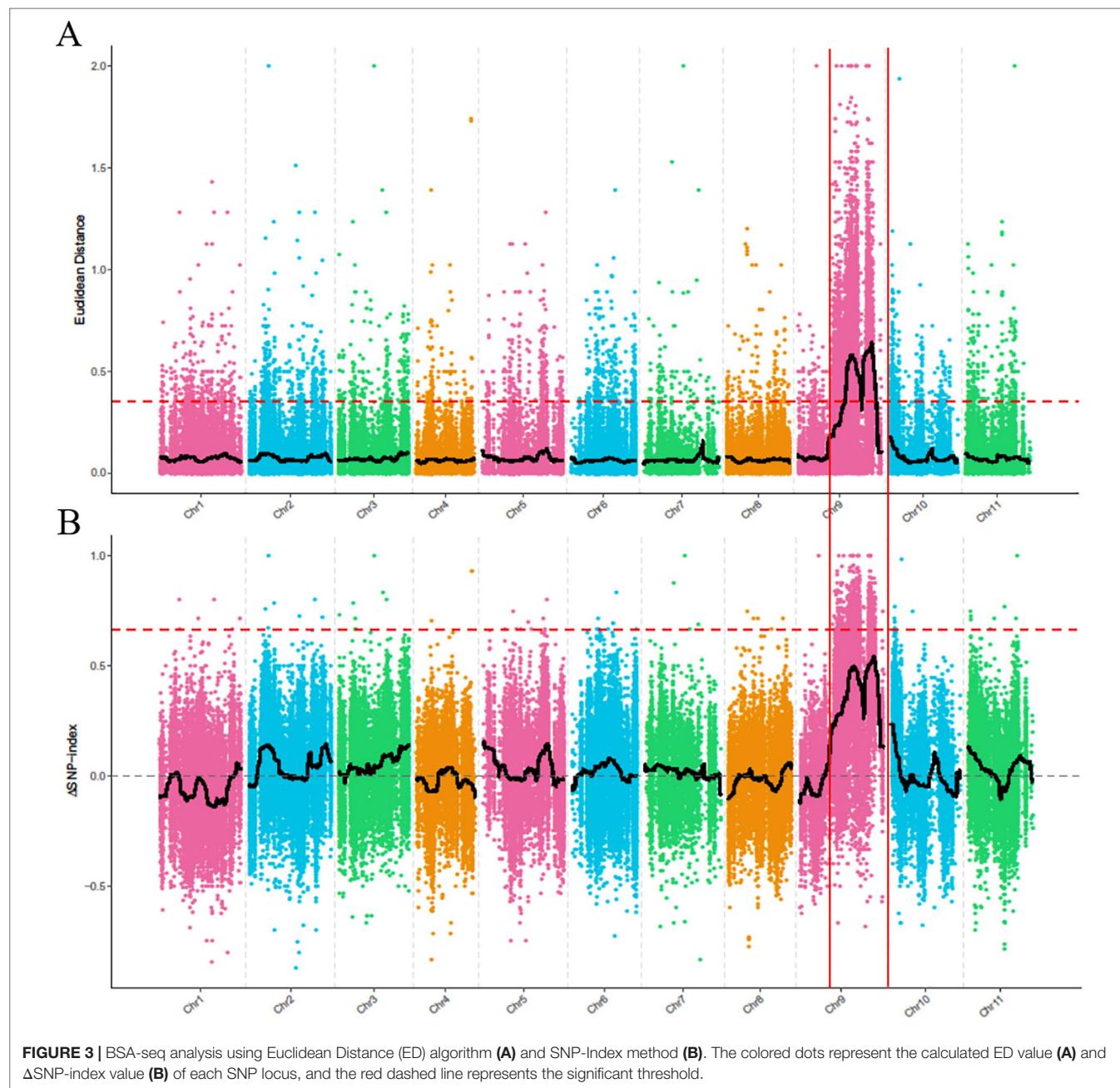
FIGURE 3 | BSA-seq analysis using Euclidean Distance (ED) algorithm **(A)** and SNP-Index method **(B)**. The colored dots represent the calculated ED value **(A)** and ΔSNP-index value **(B)** of each SNP locus, and the red dashed line represents the significant threshold.

positions. The latter two markers, ClSSR26018 and ClSSR26045, identified from the larger mapping population, were located on the same scaffold with a physical interval of 170.05 kb. For fine mapping of the *Cldw-1* gene, these two closely linked markers (ClSSR26018 and ClSSR26045) were used to genotype an extended, even larger F_2 mapping population of 1053 plants. Additionally, the whole genome re-sequences of the two parental lines were aligned to the watermelon reference genome, and SNPs and Indels were identified in this candidate region for marker development. Three dCAPS markers and one Indel marker were developed for genotyping the recombinant plants. Finally, the *Cldw-1* gene was mapped between dCAPS1 and ClSSR26045, which had 14 and 2 recombinants,

respectively. They were physically located in a 107 kb region from 30,394,615 to 30,501,700 kb on chromosome 9. Of the new mapped markers, dCAPS2, dCAPS3, and Indel1 were all co-segregated with the *Cldw-1* locus (**Figure 4C**).

According to the annotation of the watermelon reference genome 97103, there were six genes predicted to be in the 107 kb candidate region, namely *Cla010332*, *Cla010333*, *Cla010334*, *Cla010335*, *Cla010336*, and *Cla010337*. The physical positions and gene annotations of these genes are provided in **Table S3**. The SNPs and Indels between the two parental lines were checked with reference to these genes, and no difference was detected for five of the six genes (*Cla010332* to *Cla010336*). For *Cla010337*,

FIGURE 4 | Fine mapping of *Cldw-1* in watermelon. **(A)** Validation of the BSA-seq result by linkage analysis of the *Cldw-1* gene with SSR markers in 124 F$_2$ plants. **(B)** Fine mapping of the *Cldw-1* gene using eight closely flanked SSR markers in 414 F$_2$ plants. **(C)** The *Cldw-1* gene was finally mapped between marker dCAPS1 and ClSSR26045 in 1,053 F$_2$ plants. The numbers above chromosomes represented the recombinants between adjacent markers in the 1,053 F$_2$ plants.

three SNPs and two Indels were detected between the two parental lines. For the three SNPs, one SNP was located in the intron, and two SNPs were located in the exons (**Figure 5**). Of the two Indels, a 1-bp deletion was present in the 4th exon of the dwarf line, WM102, with another 5-bp deletion being detected in the 8th intron of WM102, and two markers, dCAPS3 and Indel1, were developed from these two Indels, respectively. Both of them showed co-segregation with the dwarf trait in the F$_2$ population, indicating that *Cla010337* was the candidate gene for *Cldw-1*.

We compared the expression levels of *Cla010337* in the root, hypocotyl, leaf, and stem tissues of WT4 and WM102 using qRT-PCR. The results showed that *Cla010337* was expressed in all the tissues tested, with the highest expression level being in the stem

(**Figure 6**). The expression levels in each of the test tissues from the dwarf line, WM102, were significantly lower than those from the normal-height line, WT4, suggesting that the 1-bp deletion frameshift mutation in the 4th exon greatly affected the expression and function of *Cla010337*.

Gene Annotation, Sequence Alignment, and Allelic Diversity of *Cldw-1*

The gene size of *Cla010337* in the watermelon reference genome 97103 was 6,377 bp, which was predicted to contain nine introns (**Figure S2**), and the full length of the coding sequence (CDS) was 3,753 bp, encoding a protein of 1,250 amino acids.

FIGURE 5 | Gene prediction, sequence comparison, and motif identification of the Cldw-1 gene in watermelon. **(A)** The genomic structure of the *Cldw-1* gene in watermelon. Two SNPs and one nucleotide deletion in the exons and a 5-bp deletion in the intron (marker Indel1) were detected between the two parental lines; their physical positions in the genomic sequence are indicated. The left nucleotides and the right nucleotides at each site represent the normal-height line, WT4, and the dwarf line, WM102, respectively. **(B)** The *Cldw-1* gene was annotated to encode four conserved domains in the deduced amino acid sequence.

FIGURE 6 | Expression analysis of the *Cldw-1* gene in different tissues between WM102 and WT4 using qRT-PCR. Expression of *Cldw-1* was decreased in all WM102 tissues. Error bars represent the SE. Significant differences were determined according to Student's t-test (**$P < 0.01$).

The single nucleotide deletion of the 631th in the CDS resulted in a frameshift mutation in the dwarf line, WM102, which led to a truncated protein (**Figure 5A**). Gene prediction and function annotation revealed that *Cla010337* encoded an ABC transporter that belongs to the ATP-binding cassette transporter superfamily. The deduced amino acid sequence of Cla010337 contained two conserved ABC_membrane domains and two conserved ATP-binding domains (**Figure 5B**, **Figure S3**). The ABC transporter superfamily is divided into eight subfamilies

(A-H), with Cla010337 being a member of the B subfamily. We further constructed a phylogenetic tree using ClDW-1 and other subfamily B proteins, including all the ABCB subfamily members of *Arabidopsis*, and another two ABCB proteins, ZmBR2 and SbDW3, which regulated the height of dwarf plants in maize and sorghum, respectively (Multani et al., 2003; Ye et al., 2013). The phylogenetic analysis revealed that ClDW-1 showed a close relationship with a group of well-known proteins regulating plant height, comprising AtABCB1, ZmBR2, and SbDW3 (**Figure 7**). This suggested that Cldw-1 in watermelon may have similar functions in regulating plant height development.

The allelic diversity of the *Cldw-1* gene in natural watermelon germplasm was examined by observing the plant height of other 165 watermelon accessions (**Table S1**) and through scoring in 2017 and 2018 field trials; normal plant height was exhibited in all cases. Both the Indel1 and dCAPS3 markers were then used for genotyping this worldwide collection of accessions. The results showed that only the dCAPS3 marker was completely consistent with the phenotyping (**Figure S4A**), while the Indel1 marker developed from the 8th intron was not in complete agreement with the phenotype (**Figure S4B**), suggesting that the 5-bp deletion in the intron of *Cla010337* was not conserved among different watermelon accessions.

DISCUSSION

Plant height is an important trait that affects crop architecture, resistance to lodging, fruit yield, and mechanical harvesting. The
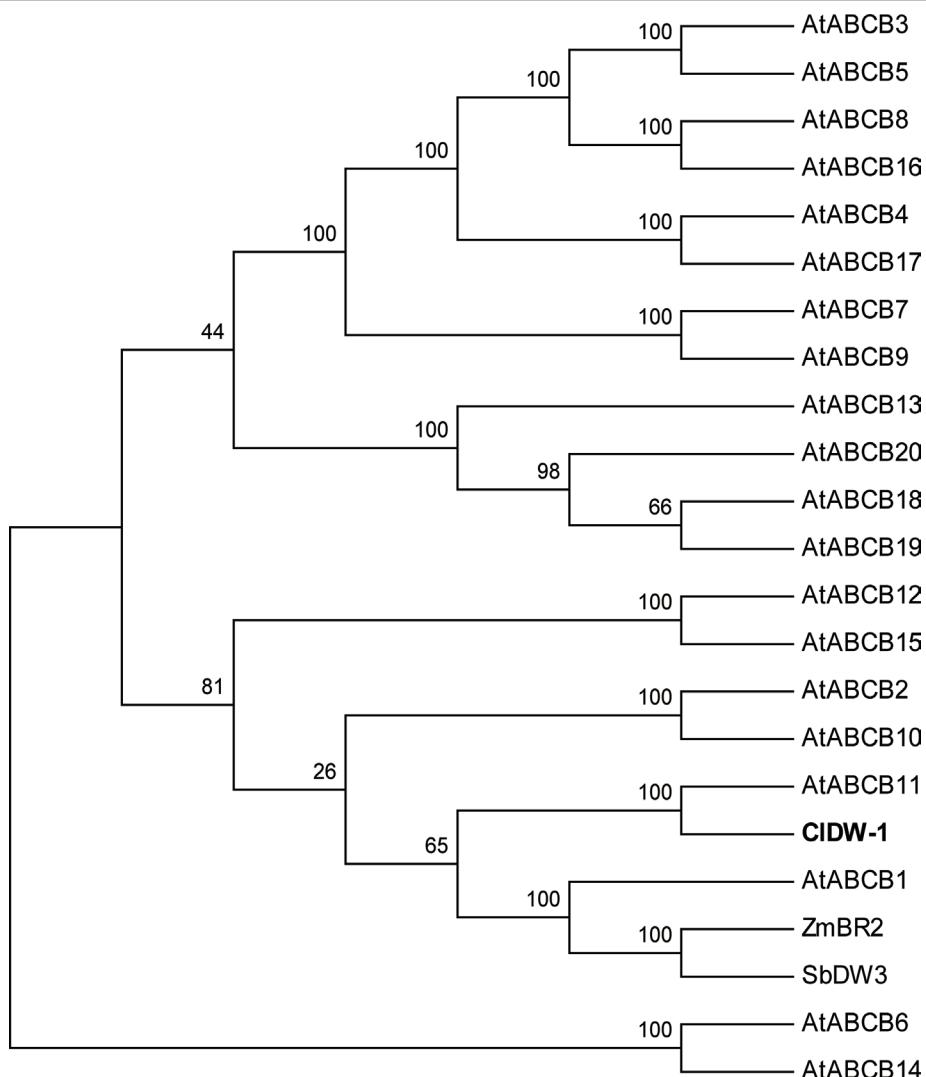
FIGURE 7 | Phylogenetic analysis of proteins of ABC transporter subfamily B. The proteins used for phylogenetic analysis including 20 genes from *Arabidopsis* (AtABCB1-19), one from maize (ZmBR2), one from sorghum (SbDW3), and one from watermelon (ClDW-1).

two main cultivation methods for watermelon in China are in the open field (creeping habit) and in protected cultivation (hanging vine habit). Most of the commercial cultivars of watermelon have tall stems with long branches, and a lot of time and labor is expected to carry out pruning. A dwarf watermelon plant with short internodes of the main stem and branches can increase the plant density and hence enhance yield per unit land area and reduce labor requirements for pruning. As a consequence, the development of dwarf or compact watermelon cultivars has become one of the main targets in watermelon breeding.

Dwarf genetic resources are important for genetic improvement and crop breeding for altered plant architecture, and at least four recessive dwarf genes have been identified in watermelon (Guner and Wehner, 2004). The WM102 used in this study was developed from 'Bush Sugar Baby,' carrying the *dw-1* gene (Mohr, 1956), and it had short internodes with two to three

branches. Cytological comparison between normal and dwarf plants in the present study revealed that cell size was significantly smaller in WM102, suggesting that the loss-of-function of *Cldw-1* led to the short cells and internodes (**Figure 2**). *dw-1ˢ* is allelic to the *dw-1* gene, with plants expressing *dw-1ˢ* having vine lengths intermediate between normal height and dwarf (Dyutin and Afanas'eva, 1987). The *dw-2* plants have a main stem length of approximately 100 cm and internode lengths of approximately 5 cm, with 5-11 branches (Robinson et al., 1976), while plants with *dw-3* have dwarf stems with fewer, lobed leaves (Huang et al., 1998). A dwarf plant was recently identified with small fruit, and genetic analysis revealed that it was controlled by a recessive gene, *dsh* (Li et al., 2016). Though some of these dwarfing genes in watermelon have been known for a long time, none of them have been cloned, and how they control plant height is still unknown.

Compared with the reverse-genetics approach in gene cloning, map-based cloning is labor-intensive and time-consuming, which restricts its use in gene identification in horticultural crops. However, forward genetics relying on random mutagenesis can precisely identify a novel gene that acts in a particular pathway, revealing novel functions for known genes. This can provide a true understanding of gene function and the mechanism by which gene networks regulate a target trait (Gillmor et al., 2016; Yang et al., 2018). In comparison with model species, only a few genes have been identified by map-based cloning in watermelon. With the high efficiency and cost-effectiveness of next-generation sequencing technology, BSA-seq can quickly detect thousands of markers across the genome with adequate coverage and identify the candidate regions associated with the target trait. This approach has been used not only in gene/QTL mapping (Lu et al., 2014; Song et al., 2017; Knorst et al., 2019; Liu et al., 2019) but also for sequencing and genome-wide association studies (Turner et al., 2010; Yang et al., 2015; Zou et al., 2016). The power of BSA-seq is affected by the size of the bulked sample of extreme individuals, sequencing strategy and depth, and the genetic architecture of the target trait (Zou et al., 2016). In the present study, we used BSA-seq for primary mapping by bulk samples of 20 dwarf plants and 20 normal height plants with genome coverage of 34 × for each bulk, and two regions on chromosome 9 were identified as candidate intervals for the *Cldw-1* gene (**Figure 3**). The identified candidate regions were discontinuous and large, which probably resulted from the relatively small sample size and the lower coverage of each bulk. However, it still provided the correct direction for further fine mapping of the *Cldw-1* gene by linkage analysis, and it was efficient in the initial mapping of a target trait. Furthermore, the marker dCAP3 developed in this study co-segregated with the dwarf trait in the F_2 mapping population and the natural population (**Figure S4**), indicating that the functional dCAP3 marker of the *Cldw-1* gene could be used in marker-assisted selection (MAS) for plant height breeding or as the basis for genome selection breeding in watermelon.

Plant hormones play central roles in the regulation of plant height by coordinating cell elongation and division, and the dwarf plant trait is controlled by genes that are involved in a complex regulatory network responsible for the biosynthesis or signal transduction of gibberellins (GAs) and brassinosteroids (BRs) (Liu et al., 2018). In the Cucurbitaceae family, several dwarf or compact habit genes associated with the GA or BR pathway have been mapped or cloned. In cucumber, the dwarf or compact genes *scp-1* and *scp-2* have been shown to be members of the plant cytochrome P450 monooxygenase gene *CsCYP85A1* and *CsDET2*, and both have been shown to have functions in the BR biosynthesis pathway (Hou et al., 2017; Wang et al., 2017). The *gibberellin* (*GA*) *20-oxidase-like* gene was identified as the candidate gene for the *dsh* gene controlling the dwarf phenotype in watermelon (Dong et al., 2018) and a major QTL for dwarf bush type in pumpkin (Zhang et al., 2015). Furthermore, GA-response or -biosynthesis mutants have been extensively used to reduce plant height in wheat, maize, and rice. Except for GAs and

BRs, dwarfing genes involved in the auxin pathways have also been discovered. Auxins have been identified as being involved in modulating plant height in maize, sorghum, and *Arabidopsis* by creating a gradient regulated *via* its transporters, such as members of the ABCB subfamily (Noh et al., 2001; Multani et al., 2003; Ye et al., 2013). The dwarfing mutations in maize (*ZmBR2*) and sorghum (*SbDW3*) reduced plant height by compacting the lower internodes without any adverse effect on the remainder of the plant (Multani et al., 2003). In *Arabidopsis*, the mutant of its ortholog, *AtPGP1-2*, and the *atabcb1 atabcb19* double knockout exhibited reduced plant height (Noh et al., 2001; Ye et al., 2013). The *Cldw-1* dwarfing gene identified in the current study also encoded an ABC transporter of the ABCB subfamily and clustered with AtABCB1, ZmBR2, and SbDW3 in the phylogenetic analysis, suggesting that it may have a conserved function in regulating plant height in both monocots and dicots. Future research will focus on the function and regulatory network of the *Cldw-1* gene, which will be helpful for elucidating the molecular mechanism of plant height regulation in watermelon.

DATA AVAILABILITY STATEMENT

GenBank accession for the re-sequencing dataset of WT4 and WM102 (PRJNA551784).

AUTHOR CONTRIBUTIONS

MZ, SS, JL, HL, SY, HY, and KZ performed phenotyping in F_2 plants and fine mapping. LY, HZ, DL, and JH contributed to data processing and analysis. SY contributed to microscopic analysis. LY and HZ wrote the manuscript. All authors reviewed and approved this manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01399/full#supplementary-material

TABLE S1 | The list of watermelon germplasm used for analyzing allelic diversity of Cldw-1 gene.

TABLE S2 | The list of primers used for mapping and qRT-PCR analysis.

TABLE S3 | The physical position and gene annotation for the six genes in the candidate region.

FIGURE S1 | Comparison of hypocotyl length, plant height and internode length between two parental lines. The plants measured were grown in the greenhouse in 2018 Spring. **(A)** The hypocotyl length of 15-day-old seeding. The plant height **(B)** and average internode length © of 30-day-old and 60-day-old plants.

FIGURE S2 | The genomic sequence and exon-intron structure of ClDW-1 gene. The highlight sequences were indicated as exons.

FIGURE S3 | Sequence alignment of ClDW-1 with its homologs in other species for ABC_membrane domain and ATP-binding domain. The

proteins used for sequence alignment including 20 genes from *Arabidopsis* (AtABCB1-19), one from maize (ZmBR2), one from sorghum (SbDW3), and one from watermelon (ClDW-1).

FIGURE S4 | The allelic diversity of Cldw-1 gene in different watermelon germplasms using marker dCAPS3 **(A)** and Indel1 **(B)**. M represented marker (100bp DNA ladder), P1 represented the normal height line WT4, P2 represented the dwarf line WM102, the number from 1 to 165 was corresponded to the natural watermelon germplasms in accordance with the code in **Table S1**.

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

Branham, S. E., Levi, A., Farnham, M. W., and Wechter, P. W. (2017). A GBS-SNP-based linkage map and quantitative trait loci (QTL) associated with resistance to *fusarium oxysporum* f. sp. niveum race 2 identified in *Citrullus lanatus* var. *citroides. Theor. Appl. Genet.* 130 (2), 319–330. doi: 10.1007/s00122-016-2813-0

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 6 (9), 677–681. doi: 10.1038/nmeth.1363

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic. Acids Res.* 31, 3497–3500. doi: 10.1093/nar/gkg500

Denna, D. W. (1962). A study of the genetic, morphological and physiological basis for the bush and vine habit of several cucurbits. [dissertation]. [Ithaca (NY)]: Cornell University.

de Saint Germain, A., Ligerot, Y., Dun, E. A., Pillot, J. P., Ross, J. J., Beveridge, C. A., et al. (2013). Strigolactones stimulate internode elongation independently of gibberellins. *Plant Physiol.* 163 (2), 1012–1025. doi: 10.1104/pp.113.220541

Do, T. H. T., Martinoia, E., and Lee, Y. (2018). Functions of ABC transporters in plant growth and development. *Curr. Opin. Plant Biol.* 41, 32–38. doi: 10.1016/j.pbi.2017.08.003

Dong, W., Wu, D., Li, G., Wu, D., and Wang, Z. (2018). Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci. Rep-UK.* 8, 2908. doi: 10.1038/s41598-018-21293-1

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 (4), 947–958. doi: 10.1007/s00122-018-3050-5

Dyutin, K. E., and Afanas'eva, E. A. (1987). Interaction of short stem in watermelon. *Tsitol. Genet.* 21 (3), 227–229.

Gillmor, C. S., Roeder, A. H., Sieber, P., Somerville, C., and Lukowitz, W. (2016). A genetic screen for Mmutations affecting cell division in the *Arabidopsis thaliana* embryo identifies seven loci required for cytokinesis. *PloS One* 11 (1), e0146492. doi: 10.1371/journal.pone.0146492

Guner, N., and Wehner, T. C. (2004). The genes of watermelon. *Hortscience.* 39 (6), 1175–1182. doi: 10.21273/HORTSCI.39.6.1175

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2012). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45 (1), 51–60. doi: 10.1038/ng.2470

Hil, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., Su, Y. C., and Yost, H. J. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23 (4), 687–697. doi: 10.1101/gr.146936.112

Hong, Z., Ueguchi-Tanaka, M., and Matsuoka, M. (2004). Brassinosteroids and rice architecture. *J. Pestic. Sci.* 29 (3), 184–188. doi: 10.1584/jpestics.29.184

Hou, S., Niu, H., Tao, Q., Wang, S., Gong, Z., Li, S., et al. (2017). A mutant in the CsDET2 gene leads to a systemic brassinosteriod deficiency and super compact phenotype in cucumber (*Cucumis sativus* L.). *Theor. Appl. Genet.* 130 (8), 1693–1703. doi: 10.1007/s00122-017-2919-z

Huang, H., Zhang, X., Wei, Z., Li, Q., and Li, X. (1998). Inheritance of male-sterility and dwarfism in watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai]. *Sci. Hortic-Amsterdam.* 74, 175–181. doi: 10.1016/S0304-4238(97)00102-7

Hwang, J., Oh, J., Kim, Z., Staub, J. E., Chung, S. M., and Park, Y. (2014). Fine genetic mapping of a locus controlling short internode length in melon (*Cucumis melo* L.). *Mol. Breeding.* 34 (3), 949–961. doi: 10.1007/s11032-014-0088-1

Hwang, J. U., Song, W. Y., Hong, D., Ko, D., Yamaoka, Y., Jang, S., et al. (2016). Plant ABC transporters enable many unique aspects of a terrestrial plant's lifestyle. *Mol. Plant* 9 (3), 338–355. doi: 10.1016/j.molp.2016.02.003

Kaneda, M., Schuetz, M., Lin, B. S., Chanis, C., Hamberger, B., Western, T. L., et al. (2011). ABC transporters coordinately expressed during lignification of *Arabidopsis* stems include a set of ABCBs associated with auxin transport. *J. Exp. Bot.* 62 (6), 2063–2077. doi: 10.1093/jxb/erq416

Kauffman, C. S., and Lower, R. L. (1976). Inheritance of an extreme dwarf plant type in the cucumber. *J. Am. Soc Hortic. Sci.* 101, 150–151.

Kim, K. H., Hwang, J. H., Han, D. Y., Park, M., Kim, S., Choi, D., et al. (2015). Major Quantitative Trait Loci and Putative Candidate Genes for Powdery Mildew Resistance and Fruit-Related Traits Revealed by an Intraspecific Genetic Map for Watermelon (*Citrullus lanatus* var. *lanatus*). *PloS One* 10 (12), e0145665. doi: 10.1371/journal.pone.0145665

Knavel, D. E. (1990). Inheritance of a short-internode mutant of 'Mainstream' muskmelon. *Hortscience.* 25, 1274–1275. doi: 10.21273/HORTSCI.25.10.1274

Knorst, V., Byrne, S., Yates, S., Asp, T., Widmer, F., Studer, B., et al. (2019). Pooled DNA sequencing to identify SNPs associated with a major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Theor. Appl. Genet.* 132 (4), 947–958. doi: 10.1007/s00122-018-3250-z

Kong, Q., Yuan, J., Gao, L., Zhao, S., Jiang, W., Huang, Y., et al. (2014). Identification of suitable reference genes for gene expression normalization in qRT-PCR analysis in watermelon. *PloS One* 9 (2), e90612. doi: 10.1371/journal.pone.0090612

Kubes, M., Yang, H., Richter, G. L., Cheng, Y., Mlodzinska, E., Wang, X., et al. (2012). The Arabidopsis concentration-dependent influx/efflux transporter ABCB4 regulates cellular auxin levels in the root epidermis. *Plant J.* 69 (4), 640–654. doi: 10.1111/j.1365-313X.2011.04818.x

Kubicki, B., Soltysiak, U., and Korzeniewska, A. (1986). Induced mutation in cucumber (Cucumis sativus L.) V. Compact type of growth. *Genet. Pol.* 27, 289–298. doi: 10.1186/s12870-015-0635-x

Lambel, S., Lanini, B., Vivoda, E., Fauve, J., Patrick Wechter, W., Harris-Shultz, K. R., et al. (2014). A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. *Theor. Appl. Genet.* 127 (10), 2105–2115. doi: 10.1007/s00122-014-2363-2

Li, Y., Yang, L., Pathak, M., Li, D., He, X., and Weng, Y. (2011). Fine genetic mapping of cp: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L. *Theor. Appl. Genet.* 123, 973-983. doi: 10.1007/s00122-011-1640-6

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics.* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y., Xu, A., Dong, W., Li, Z., and Li, G. (2016). Genetic analysis of a dwarf vine and small fruit watermelon mutant. *Hortic. Plant J.* 2 (4), 224–228. doi: 10.1016/j.hpj.2016.10.002

Liu, F., Wang, P. D., Zhang, X. B., Li, X. F., Yan, X. H., Fu, D. H., et al. (2018). The genetic and molecular basis of crop height based on a rice model. *Planta* 247 (1), 1–26. doi: 10.1007/s00425-017-2798-1

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

Liu, P. B. W., and Loy, J. B. (1972). Inheritance and morphology of two dwarf mutants in watermelon. *J. Am. Soc Hortic. Sci.* 97, 745–748.

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near *Flowering Locus* T in cucumber. *Theor. Appl. Genet.* 127 (7), 1491–1499. doi: 10.1007/s00122-014-2313-z

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi: 10.1101/gr.107524.110

Mohr, H. C. (1956). Mode of inheritance of bushy growth characteristic in watermelon. *Proc. Assoc. South Agric. Work.* 53, 174.

Mohr, H. C., and Sandhu, M. S. (1975). Inheritance and morphological traits of a double recessive dwarf in watermelon, *Citrullus lanatus* (Thunb.) Mansf. *J. Am. Soc Hortic. Sci.* 100, 135–137.

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19), 4321–4326. doi: 10.1093/nar/8.19.4321

Multani, D. S., Briggs, S. P., Chamberlin, M. A., Blakeslee, J. J., Murphy, A. S., and Johal, G. S. (2003). Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. *Science.* 302 (5642), 81–84. doi: 10.1126/science.1086072

Neff, M. M., Neff, J. D., Chory, J., and Pepper, A. E. (1998). dCAPS, a simple technique for the genetic analysis of singlenucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J.* 14 (3), 387–392. doi: 10.1046/j.1365-313X.1998.00124.x

Niemirowicz-Szczytt, K., Rucinska, M., and Korzeniewsia, A. (1996). An induced mutation in cucumber (Cucumis sativus L.): super compact. *Cucurbit Genet. Coop. Rep.* 19, 1–3.

Noh, B., Murphy, A. S., and Spalding, E. P. (2001). Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development. *Plant Cell.* 13 (11), 2441–2454. doi: 10.1105/tpc.010350

Nomura, T., Jager, C. E., Kitasaka, Y., Takeuchi, K., Fukami, M., Yoneyama, K., et al. (2004). Brassinosteroid deficiency due to truncated steroid 5alpha-reductase causes dwarfism in the *lk* mutant of pea. *Plant Physiol.* 135 (4), 2220–2229. doi: 10.1104/pp.104.043786

Paris, H. S., Nerson, H., and Karchi, Z. (1984). Genelics of internode length in melons. *J. Hered.* 75 (5), 403–406. doi: 10.1093/oxfordjournals.jhered.a109965

Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., et al. (2011). Molecular characterization of *Rht-1* dwarfing genes in hexaploid wheat. *Plant Physiol.* 157 (4), 1820–1831. doi: 10.1104/pp.111.183657

Pitrat M. (1991). Linkage groups in *Cucumis melo* L. *J. Hered.* 82 (5), 406–411. doi: 10.1093/oxfordjournals.jhered.a111112

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 (1), 836–850. doi: 10.1104/pp.17.01290

Ren, Y., Jiao, D., Gong, G. Y., Zhang, H. Y., Guo, S. G., Zhang, J., et al. (2015). Genetic analysis and chromosome mapping of resistance to *Fusarium oxysporum* f. sp niveum (FON) race 1 and race 2 in watermelon (*Citrullus lanatus* L.). *Mol. Breeding.* 35, 183. doi: 10.1007/s11032-015-0375-5

Ren, Y., Zhao, H., Kou, Q., Jiang, J., Guo, S., Zhang, H., et al. (2012). A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PloS One* 7 (1), e29453. doi: 10.1371/journal.pone.0029453

Robinson, R. W., Munger, H. M., and Whitaker, T. W. (1976). Genes of Cucubitaceae. *Hortscience.* 11, 554–568.

Sakamoto, T., Miura, K., Itoh, H., Tatsumi, T., Ueguchi-Tanaka, M., Ishiyama, K., et al. (2004). An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol.* 134 (4), 1642–1653. doi: 10.1104/pp.103.033696

Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., et al. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature.* 416 (6882), 701–702. doi: 10.1038/416701a

Song, J., Li, Z., Liu, Z., Guo, Y., and Qiu, L. J. (2017). Next-Generation Sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Front. Plant Sci.* 8, 919. doi: 10.3389/fpls.2017.00919

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74 (1), 174–183. doi: 10.1111/tpj.12105

Tamiru, M., Undan, J. R., Takagi, H., Abe, A., Yoshida, K., Undan, J. Q., et al. (2015). A cytochrome P450, OsDSS1, is involved in growth and drought stress responses in rice (*Oryza sativa* L.). *Plant Mol. Biol.* 88 (1–2), 85–99. doi: 10.1007/s11103-015-0310-5

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., and Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42 (3), 260–263. doi: 10.1038/ng.515

Wang, H., Li, W., Qin, Y., Pan, Y., Wang, X., Weng, Y., et al. (2017). 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, 266. doi: 10.3389/fpls.2017.00266

Wei, C., Chen, X., Wang, Z., Liu, Q., Li, H., Zhang, Y., et al. (2017). Genetic mapping of the *LOBED LEAF 1* (*ClLL1*) gene to a 127.6-kb region in watermelon (*Citrullus lanatus* L.). *PloS One* 12 (7), e0180741. doi: 10.1371/journal.pone.0180741

Yang, J. L., Jiang, H. Y., Yeh, C. T., Yu, J. M., Jeddeloh, J. A., Nettleton, D., et al. (2015). Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant J.* 84 (3), 587–596. doi: 10.1111/tpj.13029

Yang, L., Liu, H., Zhao, J., Pan, Y., Cheng, S., Lietzow, C. D., et al. (2018). *LITTLELEAF* (*LL*) encodes a WD40 repeat domain-containing protein associated with organ size variation in cucumber. *Plant J.* 95, 834–847. doi: 10.1111/tpj.13991

Yang, S., Wen, C. L., Liu, B., Cai, Y. L., Xue, S. D., Bartholomew, E. S., et al. (2019). A *CsTu-TS1* regulatory module promotes fruit tubercule formation in cucumber. *Plant Biotechnol. J.* 17 (1), 289–301. doi: 10.1111/pbi.12977

Ye, L. F., Liu, L., Xing, A. Q., and Kang, D. M. (2013). Characterization of a dwarf mutant allele of *Arabidopsis* MDR-like ABC transporter AtPGP1 gene. *Biochem. Bioph. Res. Co.* 441 (4), 782–786. doi: 10.1016/j.bbrc.2013.10.136

Zhang, G., Ren, Y., Sun, H., Guo, S., Zhang, F., Zhang, J., et al. (2015). A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). *BMC Genomics* 16, 1101. doi: 10.1186/s12864-015-2312-8

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

Zhu, H., Song, P., Koo, D. H., Guo, L., Li, Y., Sun, S., et al. (2016). Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genomics* 17, 557. doi: 10.1186/s12864-016-2870-4

Zou, C., Wang, P. X., and Xu, Y. B. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* 14 (10), 1941–1955. doi: 10.1111/pbi.12559

QTL and Transcriptomic Analyses Implicate Cuticle Transcription Factor *SHINE* as a Source of Natural Variation for Epidermal Traits in Cucumber Fruit

Stephanie Rett-Cadman[1], Marivi Colle[1], Ben Mansfeld[1], Cornelius S. Barry[1], Yuhui Wang[2,3], Yiqun Weng[2,3], Lei Gao[4], Zhangjun Fei[4] and Rebecca Grumet[1]*

[1] Department of Horticulture and Graduate Program in Plant Breeding, Genetics and Biotechnology, Michigan State University, East Lansing, MI, United States, [2] Department of Horticulture, University of Wisconsin, Madison, WI, United States, [3] USDA-ARS, Vegetable Crops Research Unit, Madison, WI, United States, [4] Boyce Thompson Institute, Cornell University, Ithaca, NY, United States

The fruit surface is a unique tissue with multiple roles influencing fruit development, post-harvest storage and quality, and consumer acceptability. Serving as the first line of protection against herbivores, pathogens, and abiotic stress, the surface can vary markedly among species, cultivars within species, and developmental stage. In this study we explore developmental changes and natural variation of cucumber (*Cucumis sativus* L.) fruit surface properties using two cucumber lines which vary greatly for these traits and for which draft genomes and a single nucleotide polymorphism (SNP) array are available: Chinese fresh market type, Chinese Long '9930' (CL9930), and pickling type, 'Gy14'. Thin-section samples were prepared from the mid-region of fruit harvested at 0, 4, 8, 12, 16, 20, 24 and 30 days post pollination (dpp), stained with Sudan IV and evaluated for cuticle thickness, depth of wax intercalation between epidermal cells, epidermal cell size and shape, and number and size of lipid droplets. 'Gy14' is characterized by columnar shaped epidermal cells, a 2–3 fold thicker cuticular layer than CL9930, increased cuticular intercalations between cells and a larger number and larger sized lipid droplets. In both lines maximal deposition of cuticle and increase in epidermal size coincided with exponential fruit growth and was largely completed by approximately 16 dpp. Phenotyping and quantitative trait locus mapping (QTL) of fruit sampled from an F_7:F_8 Gy14 × CL9930 recombinant inbred line (RIL) population identified QTL regions on chromosomes 1, 4 and 5. Strong QTL for epidermal cell height, cuticle thickness, intercalation depth, and diameter of lipid droplets co-localized on chromosome 1. SSR markers on chromosome 1 were used to screen for recombinants in an extended RIL population to refine the QTL region. Further fine mapping by KASP assay combined with gene expression profiling suggested a small number of candidate genes. Tissue specificity, developmental analysis of expression, allelic diversity and gene function implicate the regulatory factor *CsSHINE1/WIN1* as a source of natural variation for cucumber fruit epidermal traits.

Keywords: cucumber, *Cucumis sativus*, fruit surface, cuticle, lipid droplet, epidermis, fruit development

INTRODUCTION

Fruit surfaces play important roles in fruit development, maturation, and post-harvest quality. During growth and maturation, the exocarp is the first line of defense against herbivores, pathogens, and abiotic stresses such as dehydration, UV irradiation, and mechanical pressure. Following harvest, morphological features of the fruit surface can influence consumer preference and fruit quality. The epidermal cell structure can influence fruit firmness and susceptibility to damage. Cuticle structure and waxiness can modify external appearance such as glossiness and uniformity, and modulate rate of evaporative water loss, susceptibility to cracking and pathogen infection, and material penetration into the fruit surface (reviews: Hen-Avivi et al., 2014; Lara et al., 2014; Martin and Rose, 2014). These factors, in turn, influence handling practices in the commercial market chain.

For cucumber (*Cucumis sativus*), different market types vary substantially with respect to fruit surface features that influence consumer preferences, suitability for shipping, handling and storage, or performance demands for processing (pickling). Cucumber market types can vary considerably with regard to post-harvest longevity. Weight loss during the market chain is a primary concern and is influenced by epidermal properties including skin toughness and waxiness (Patel and Panigrahi, 2019; https://www.postharvest.net.au/product-guides/cucumber/). Wax load is inversely related with rate of water loss (Wang et al., 2015a), and untreated cucumber fruit can have a shelf life of less than a week (Patel and Panigrahi, 2019). As a result, packaging methods for fresh market cucumbers include wrapping in plastic, which is both expensive and environmentally undesirable, or treating the fruit with edible coatings.

The importance of the cuticle in product quality of fleshy fruits, coupled with relative ease of isolation from certain species, has driven studies focused on the biosynthesis and properties of fruit cuticles (Hen-Avivi et al., 2014; Lara et al., 2014; Martin and Rose, 2014). While much of the cuticle biosynthetic pathway has been established in Arabidopsis, characterization of mutants with altered cuticle composition (Isaacson et al., 2009; Nadakuduti et al., 2012; Yeats et al., 2012; Petit et al., 2014), tissue-specific transcriptomic analysis of developing tomato fruit peel (Mintz-Oron et al., 2008; Matas et al., 2011), and QTL mapping of introgression lines of wild tomato species (*Solanum pennelli*, *Solanum habroachaites*) (Cohen et al., 2017; Fernandez-Moreno et al., 2017) also have identified numerous genes associated with fruit cuticle development and composition. In cucumber, homologs of two key cuticle biosynthetic enzyme genes involved in cutin and wax biosynthesis, eceriferum (*CER1*) and *WAX2*, have been cloned (Wang et al., 2015b). Decreased expression of *CER1* and *WAX2* was associated with reduced wax load and increased water loss from harvested fruits.

Deposition of the cuticle and epicuticular waxes is developmentally programmed during organ growth to accommodate coverage required by increased surface area (Martin and Rose, 2014; Ingram and Nawrath, 2017). The precursors needed for cuticle and wax deposition are produced by epidermal cells and delivered to the fruit surface (Kunst and Samuels, 2003; Matas et al., 2010; Matas et al., 2011; Yeats and

Rose, 2013; Huang 2018). Accordingly, *CsCER1* and *CsWAX2* are preferentially expressed in cucumber epidermal tissue (Ando et al., 2012; Wang et al., 2015a, b). For many species, cuticle deposition ceases during early fruit development, often before the fruit has reached maximum size and prior to the onset of ripening (Lara et al., 2014). We have observed that cuticle thickness in the pickling cucumber cultivar, 'Vlaspik', increases dramatically during the rapid growth phase from 4 to 16 days post-pollination (dpp) (Ando et al., 2012; Ando et al., 2015). The time period of 8–12 dpp also was marked by peak expression of genes associated with cuticle biosynthesis, such as several extracellular GDSL motif lipase/hydrolase proteins and lipid transfer proteins which have been implicated in lipid transport to extracellular surface (Ando et al., 2012).

Cuticle-related transcription factors have been identified from Arabidopsis, including the AP2 domain superfamily member, shine1 (SHN1), or win1 (WAX INDUCER1) (Aharoni et al., 2004; Broun et al., 2004). In tomato fruit, an exocarp-expressed *SHN* clade member, *SlSHN3*, regulates cuticle production; suppression of *SlSHN3* reduced cuticle production and caused a glossier fruit surface (Shi et al., 2013). In cucumber fruit, a preferentially peel-expressed homolog of *SHN1* (*CsaV31g030200*) exhibited peak transcription at 8–12 dpp, in concert with expression of the suite of cuticle biosynthesis associated genes (Ando et al., 2012). Several other transcription factors identified in tomato including MYB, MADS and homeodomain leucine zipper family members are also associated with regulation of production of cutin and wax components and cutin-localized secondary metabolites such as flavonols and terpenes (Adato et al., 2009; Isaacson et al., 2009; Gimenez et al., 2015). Many of the cuticle related transcription factors, including *AtSHN1/SlSHN3*, are also linked to epidermal cell patterning. In Arabidopsis and tomato *MIXTA-like* and leucine zipper transcription factors regulate both cuticle production and epidermal cell formation (Nadakuduti et al., 2012; Oshima et al., 2013; Lashbrooke et al., 2015a), and in maize, the glossy trait conferred by the AP2/EREBP transcription factor gene, *GL15,* influences epicuticular wax deposition, leaf hair formation, and cell shape (Moose and Sisco, 1996; Lauter et al., 2005).

While much of our understanding of epidermal cell structure and cuticle development has been derived from mutant or overexpression analyses using a limited number of model systems, little is known about the factors driving variation in natural populations (Petit et al., 2017). Current genetic resources and genomic tools can greatly facilitate our ability to identify and utilize sources of natural diversity across an increasing number of species, including cucumber. Reference genomes have been developed for representatives of two morphologically distinct cucumber market classes: the fresh market Chinese Long type, 'CL9930', and the American pickling type, 'Gy14' (Huang et al., 2009; Yang et al., 2012; Wang et al., 2018). In addition to obvious differences in fruit size and shape, CL9930 and Gy14 show markedly different epidermal and cuticle structures including amount and location of cuticle and wax deposition, number and size of lipid droplets present in epidermal cells, and size and shape of epidermal cells (Colle, 2015). In this study we sought to characterize epidermal cell growth and cuticle and wax

deposition during cucumber fruit development, and identify genomic regions and candidate genes associated with variation using recombinant inbred lines (RILs) derived from progeny of Gy14 × CL9930. Several QTL were identified, including a major QTL on chromosome 1 associated with cuticle thickness, epidermal cell height, intercalation depth, and diameter of lipid droplets. Combined fine mapping, transcriptional analysis, and allelic diversity among cucumber accessions implicated the transcription factor *CsSHINE1/WIN1* as a regulator of natural variation for cucumber fruit epidermal traits.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Plant Materials. Seed of cucumber (*C. sativus* L.) lines Gy14 (American pickling cucumber inbred line) and CL9930 (Chinese long type) were originally obtained from the University of Wisconsin and multiplied in the greenhouse. Pickling type cultivar Vlaspik was obtained from Seminis Vegetable Seed Inc, Oxnard, CA and American slicing type cultivar Poinsett 76 from Seedway, Hall NY. Our prior studies show that despite differences in size and shape, all four varieties exhibit a typical developmental pattern for cucumber with a period of cell division (~0–4 dpp), followed by exponential growth, and approaching full size at 16–20 dpp (Ando et al., 2012; Colle, 2015; Colle et al., 2017). The three American varieties all have thick cuticles, while CL9930 has a thin cuticle (**Supplementary Table 1**)

Developmental Study. Gy14 and CL9930 plants were grown in the greenhouse (Michigan State University Plant Science Greenhouse Complex, East Lansing MI) in summer 2017 in 4 L plastic pots with Suremix Perlite soil medium (Michigan Grower Product, Inc., Galesburg, MI). The plants were watered and fertilized twice daily (with 44 ppm nitrogen of Peters Professional 20-20-20 General Purpose; Scotts, Marysville, OH) using an automated drip irrigation system (Dositron model D14MZ2, Clearwater FL). Supplemental high pressure sodium lights were used to provide a 16-h photoperiod. Pest and disease control were performed according to standard management practices in the greenhouse. When the plants initiated female flower production, a single ovary from each of 48 plants per line were hand-pollinated on the same day to ensure comparable environmental conditions for all fruit during development and provide sufficient fruit for each harvest date. Only one fruit was set per plant to have consistent developmental rates for all fruits by preventing competition for resources among fruits. At each sample date (0, 4, 8, 12, 16, 20, 24, and 30 dpp) three fruits per line (biological replicates)/age were harvested. Three samples (technical replicates) derived from the midsection of each fruit were prepared for microscopy.

RIL Analyses. A Gy14 × CL9930-derived $F_{7:8}$ RIL population (Weng et al., 2015) that was previously genotyped by SNP array (Rubenstein et al., 2015) was grown in the greenhouse in Fall 2016, as described above. Each of 110 RILs and both parental lines were grown in triplicate (biological replicates) in a randomized, complete block design. Ovaries were hand pollinated and only one fruit per plant was set to minimize inter-fruit competition.

Fruits were harvested at 16 dpp and prepared as described below. An extended $F_{7:8}$ RIL population, consisting of 375 lines, was screened using SSR markers to identify recombinants for regions in chromosomes 1 and 4. Recombinant lines, parents, and reciprocal F_1s were grown in triplicate (biological replicates) in the field in Summer 2018, in a randomized, complete block design at the Michigan State University Horticulture Teaching and Research Center, East Lansing, MI. Bee-pollinated flowers were tagged at anthesis. Fruit were harvested at 20–22 dpp and two samples (technical replicates) derived from the midsection of each fruit were prepared for microscopy as described below. Pest and disease control were performed according to standard management practices under field conditions. A subset of 17 RILs recombinant in the region of interest on chromosome 1 along with both parents were grown in the greenhouse in Spring 2019 under conditions described above to provide replication in different environments. Fruit were harvested at 16–20 dpp.

RNA-Seq Experiment. Fruit from the cultivars 'Poinsett 76' and 'Gy 14' were grown under greenhouse conditions as described above. Flowers were hand pollinated, such that 8 and 16 dpp fruit were harvested on the same day. Peels from three fruit (biological replicates) were harvested for each age and genotype and immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction.

CsSHN1 expression analysis. Sixty plants of CL9930, Gy14, and Vlaspik were grown under greenhouse conditions as described above with the following modifications: supplemental lights provided an 18-h light cycle and plants were hand fertilized once a week. One or two flowers from the third to fifth node were hand-pollinated on the same day on each plant for each genotype; a single fruit was allowed to develop. Three fruits each from CL9930, Gy14 and Vlaspik were collected at anthesis, 4, 8, 12, 16, and 20 dpp.

Microscopy and Measurement of Epidermal Traits

A wedge (~1 cm³) was cut from the mid-section of each fruit and sliced to ~0.1 mm thickness by a sliding block microtome. All methods pertaining to staining with Sudan IV (Sigma-Aldrich, St. Louis MO) and subsequent washing were performed according to the methods Buda et al. (2009) with the exception of RIL experiments (Summer 2018 and Spring 2019) when samples were mounted in glycerin (Columbus Chemical Industries, Columbus WI) instead of distilled water. All samples in water were imaged by microscopy the same day; glycerin mounted samples were imaged within one week. Images for the RIL population from the Fall 2016 experiment were captured using an EVOS FL Auto imaging system (ThermoFisher Scientific; http://www.thermofisher.com) with 400× magnification and analyzed using the Nikon NIS-Elements BR imaging system. For the developmental study (Summer 2017) and the extended RIL population (Summer 2018, Spring 2019), images were obtained using a Nikon Eclipse Ni-U microscope and Nikon DS-Fi3 camera (Nikon Instruments Inc.; Melville, NY) at 600× and 200× magnification, respectively. Epidermal features were measured as shown in **Figure 1**. To allow for better comparison among samples
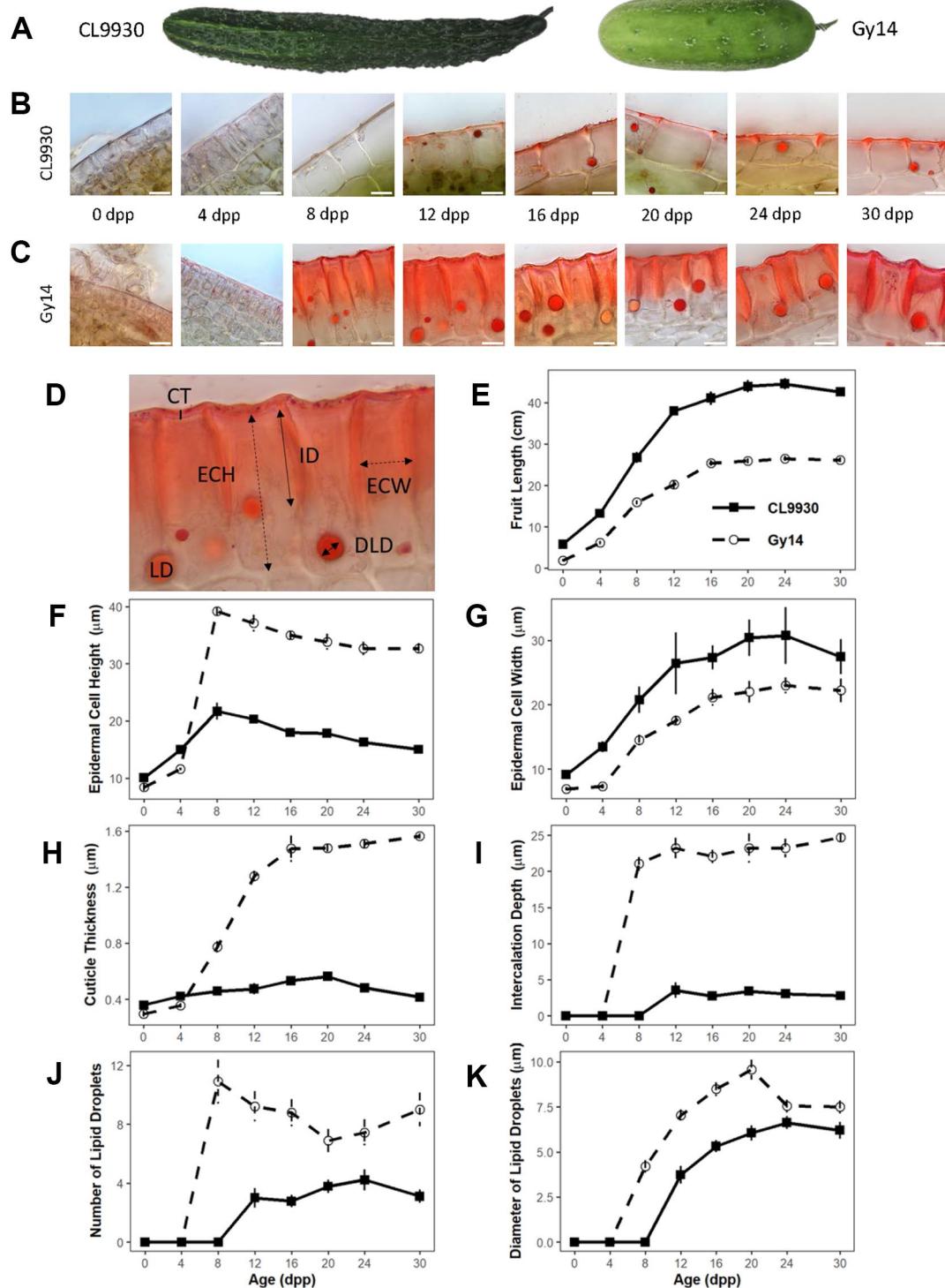
FIGURE 1 | Developmental study of cucumber fruit epidermal traits. **(A)** CL9930 and Gy14 fruit at 16 days post pollination (dpp). **(B, C)** Cross sections of CL9930 **(B)** and Gy14 **(C)** at 0, 4, 8, 12, 16, 20, 24, and 30 days post pollination (dpp). Magnification = 1,000×. Scale bar = 10 μm. Samples were taken from the mid-section of the fruit. **(D)** Cross section illustrating traits measured; lipids were stained with Sudan IV. Cuticle thickness (CT) is represented by a solid, vertical line; Epidermal cell height (radial dimension, ECH) and Epidermal cell width (ECW) by dashed, double-headed arrows; Intercalation depth (ID) by a solid, double-headed arrow; LD indicates lipid droplet, and lipid droplet diameter (LDD) is represented by a solid, double-headed arrow. **(E)** Fruit length for CL9930 (solid line) and Gy14 (dotted line). **(F–K)** Developmental progression of fruit epidermal traits for CL9930 (solid line) and Gy14 (dotted line): **(F)** epidermal cell height (radial dimension), **(G)** epidermal cell width **(H)** cuticle thickness, **(I)** intercalation depth, **(J)** number of lipid droplets in 120 μm linear region of epidermal cells, **(K)** diameter of lipid droplets. Each value is the mean of 3 replicate fruits ± S.E.

and avoid influence of warts and spines, all measurements were made in areas between spines. To standardize measurements of epidermal features, a line of 120 μm (developmental study) or 450 μm (RIL populations) was drawn across a given sample and features were measured within that area. Three measurements across the sample were taken for cuticle thickness (CT), intercalation depth (ID), and epidermal cell height (ECH); the mean value was used in subsequent analyses. Epidermal cell width (ECW) was determined by dividing 450 μm by the number of epidermal cells in that area. The number of lipid droplets (NLD) were counted in the given area. The diameters of all lipid droplets (DLD) in this area were measured and the average used in subsequent analyses. Calculation of Pearson's correlation coefficients among traits were conducted using the R package 'GGally' (https://github.com/ggobi/ggally).

Mapping of Epidermal Traits
QTL Analyses
For QTL analysis, a subset of 916 unique markers were used from a previously constructed genetic map (Rubenstein et al., 2015; Weng et al., 2015). Composite interval mapping (CIM) was performed with QTL Cartographer v2.5 using the standard model Zmapqtl 6 with walking speed of 1 cM, 5 background markers, and window size of 5 cm (Wang et al., 2012). The forward and backward method was used to select markers as cofactors. The LOD significance threshold was determined by a 1,000-permutation test at 5% probability.

SSR Screening
Microsatellite (SSR) markers within the 2.0-LOD intervals at the two QTL loci on Chr1 and Chr4 were used to genotype an expanded population of 375 $F_{7:8}$ Gy14× 9930 RILs to identify recombinants between the flanking markers. Due to the inconsistency in physical locations of the flanking SNP markers in 9930 v2.0 and Gy14 v2.0 draft genome assemblies, multiple SSR markers in the two target regions were employed. Information of markers used to identify recombinants is provided in **Supplementary Table 2**. DNA extraction, PCR amplification of molecular markers and gel electrophoreses followed Gao et al. (2016).

KASP™ Screening
DNA Isolation and Quantification. Tissue samples (~50 mg) from young leaf tissue of cucumber seedlings (~1–2 weeks) were lyophilized in a freeze-dryer and ground into fine powder with a high-throughput homogenizer (OPS Diagnostics, Lebanon, NJ). DNA was isolated and quantified as described in Wiersma et al., 2017. Briefly, DNA was isolated using the Mag-Bind® Plant DNA Plus 96 Kit (M1128, Omega Bio-Tek,Norcross, GA) on a King Fisher Flex Purification System (Thermo Scientific, Waltham, MA). DNA was quantified using the Quant-iT™ PicoGreen® dsDNA Kit (Life Technologies Corp., Grand Island, NY) on a CFX384 Real-Time thermal cycler, C1000 (BioRad, Hercules, CA).

SNP Calling and KASP™ Assay. From the expanded RIL population, 87 lines selected to be recombinants in QTL regions 1

or 4 were grown in triplicate in field conditions and phenotyped as described above. Due to the strength of the QTL on chromosome 1, further fine mapping was performed using the 29 lines that were identified by SSR assay to be recombinants in that region. The draft genomes of Gy14 (Version 2, cucurbitgenomics.org) and CL9930 (Version 3, cucurbitgenomics.org) were aligned using the *nucmer* function of MUMmer 4 (Marçais et al., 2018). Single nucleotide polymorphisms (SNPs) were then called using the *show-snps* function. SNPs immediately flanking the region of interest on chromosome 1 and at an interval of 0.25 Mb along this region were used to design allele specific forward KASP™ (LGC, Teddington, Middlesex, UK) and common reverse primers, where all CL9930 alleles would express a FAM signal, while the Gy14 allele would express a HEX signal. KASP™ markers that met the following criteria were selected for use in assay: GC content of 30–55%; approximate melting temperature of 64 ± 2°C; length of 21–28 bp; product size of 50–100 bp; limited to no secondary structure or repeats; and GC clamp with no more than 3 Gs or Cs in the last 5 bp of the primer (**Supplementary Table 3**). PCR thermocycling and fluorescence detection was conducted using a CFX384 Real-Time thermal cycler (BioRad), where alleles were determined using the CFX manager software (v.3.1).

Expression Analyses
RNA-Seq Experiment
Sample Collection and RNA Extraction. Peels were collected from 8 and 16 dpp of 'Poinsett 76' and 'Gy 14' using a vegetable peeler, immediately frozen in liquid nitrogen and stored at −80°C until further use. Three fruit (biological replicates) were collected for each age and genotype. Peel samples were ground in liquid nitrogen using a mortar and pestle. RNA extraction was performed using the MagMAX Plant RNA Isolation Kit protocol (Thermo Fisher Scientific, Waltham MA) with the exception of increased amount of tissue and buffer; approximately 100–150 mg tissue was transferred to a 1.5 ml tube with 1,000 μl of lysis buffer. After lysis and centrifugation as per the protocol, supernatant was transferred to a 96-deep-well plate for high-throughput RNA extraction, on a KingFisher Flex Purification System. Immediately after the run was complete, the 96-well plate was transferred to storage at −80°C. RNA concentration and quality were measured using Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and LabChip GX (Perkin Elmer, Waltha MA), respectively. All samples had a minimum RNA quality score of 8.

RNA-Seq Library Preparation and Sequencing. RNA-seq libraries were prepared at Michigan State University's Research Technology Support Facility, using the Illumina TruSeq Stranded mRNA Library Preparation Kit on a Sciclone G3 workstation following manufacturer's recommendations. An additional cleanup with 0.8× AmpureXP magnetic beads was performed after completion of library preparation. Quality control and quantification of completed libraries was performed using a combination of Qubit dsDNA HS and Advanced Analytical Fragment Analyzer High Sensitivity DNA assays. The libraries were divided into two pools of 15 libraries each. Pools were quantified using the Kapa Biosystems Illumina Library

Quantification qPCR kit. Each pool was loaded onto one lane of an Illumina HiSeq 4000 flow cell and sequencing was performed in a 1×50 bp single read format using HiSeq 4000 SBS reagents. Base calling was done by Illumina Real Time Analysis (RTA) v2.7.7 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

Differential Expression Analysis. Reads were cleaned, and adaptor sequences were removed using Trimmomatic v. 0.34 (Bolger et al., 2014) with the following settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:35. Quality control was performed using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). A cucumber transcriptome fasta file was made from the 'Chinese Long' (v2) (Huang et al., 2009; Li et al., 2011) genome using the *gffread* function from the cufflinks software package (Trapnell et al., 2010) and high-quality reads were then quasi-mapped to the transcriptome using Salmon v. 0.9.1 (Patro et al., 2017) with default settings.

Read quantification data was imported into R using the tximport R package (Soneson et al., 2015) and differential expression analysis was performed using DEseq2 (Love et al., 2014) with log-fold-change-shrinkage. Age and genotype were combined into a single factor for differential expression analysis and contrasts between the four conditions ('Poinsett 76' 8 dpp, 'Poinsett 76' 16 dpp, 'Gy14' 8 dpp, 'Gy14' 16 dpp) were performed. Differentially expressed genes were called significant using an adjusted p-value (Benjamini and Hochberg, 1995) of less than 0.05. A cutoff expression change of above two-fold was used to define biological significance. Expression data for candidate genes from CL9930 was accessed using the gene expression profiles function (Zheng et al., 2019) of http://cucurbitgenomics.org/; gene expression project PRJNA 312872 (Wei et al., 2016).

Expression Analysis of *CsSHN1* During Fruit Development

Pericarp samples isolated from the middle part of the fruit of CL9930, Gy14 and Vlaspik were immediately frozen in liquid nitrogen. Total RNA samples from the pericarp tissue were prepared using the Trizol method (Thermo Fisher Scientific), followed by DNase I treatment and clean up (Qiagen, Germantown MD). The amount of RNA for each sample was measured using the nanodrop ND-1000 (Thermo Fisher Scientific). First strand cDNA synthesis was performed using the High Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific) and by following the protocol described by Ando and Grumet (2010). Gene-specific primers were designed using Primer Express software (Applied Biosystems, Forest City CA). The ABI Prism 7900HT Sequence Detection System was used for qRT-PCR analysis. Revolution PCR Master Mix (Integrated Scientific Solutions, San Diego CA) with ROX as reference dye was used for gene amplification. *C. sativus polyubiquitin* (CuSa200910_13711) was used as an endogenous control for normalization. Expression of target genes was assessed with reference to corresponding standard curves. qRT-PCR was performed using cDNA of three fruits (three biological replicates)/genotype with three technical replicates/biological replicate. Data were analyzed by analysis of variance (ANOVA) and Tukey HSD protocol in SAS (SAS Institute, Cary, NC).

Analysis of *CsSHN1* Alleles

Identification of CsSHN1 Alleles in Gy14, Poinsett 76 and Vlaspik. DNA was extracted from three lines 'Gy 14', 'Poinsett 76', and 'Vlaspik' using the Kingfisher DNA extraction robot as described in Wang et al. (2018). After quantitation, all libraries were pooled in equimolar amounts which was loaded on one lane of an Illumina HiSeq 2500 High Output flow cell (v2) alongside other samples with a targeted coverage of ~30×. Sequencing was carried out using HiSeq SBS reagents in a 2×150 bp paired end format (PE150). Reads were cleaned and adaptor sequences were removed using Trimmomatic v. 0.33 (Bolger et al., 2014). Reads were mapped to the 'Chinese Long 9930' (v2) (Huang et al., 2009; Li et al., 2011) cucumber genome using BWA-MEM (Li and Durbin, 2009). Duplicate reads were marked with Picard (https://broadinstitute.github.io/picard/) and the GATK "Best Practices" pipeline was used for variant calling (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013). Variants were hard-filtered with the GATK base recommendations. Initial analyses were done with CL9930v2 but nucleotide positions were later converted to CL9930v3.

Survey of Cucumber Germplasm for CsSHN1 Alleles. Cucumber accessions for which resequencing data were available were examined for the nucleotide present at position 16961026 within the *CsSHN1* locus (*CsaV3_1G030200*). Sequence data for 115 accessions were available from Qi et al. (2013). Data for an additional 89 accessions (**Supplementary Table 4**), comprising a portion of the cucumber core outlined in Wang et al., 2018, were also analyzed. Samples were included for which there were at least 10 reads at position 16961026.

RESULTS

Developmental Progression of Fruit Epidermal Traits

The parental inbred lines Gy14 (a pickling breeding line) and CL9930 (an Asian fresh market breeding line) differ for fruit size, shape, and epidermal properties (**Figure 1**). Epidermal cells of Gy14 have a palisade orientation, thicker cuticle, and deeper cuticular intercalations between cells, whereas CL9930 has a flatter epidermal cell shape, with wider cells, thinner cuticle and minimal cuticular intercalation. An additional striking feature of the epidermal cells was the presence of large circular droplets brightly stained with the red lipid-soluble dye, Sudan IV. The number and size of the lipid droplets also differed between the two lines of interest, with larger and more numerous lipid droplets in Gy14.

A developmental study was performed in the greenhouse to assess changes in epidermal properties during fruit growth and maturation. To minimize effects of competition on growth rate, a single fruit was set per plant. Fruit were harvested at 0 (anthesis), 4, 8, 12, 16, 20, 24, and 30 dpp (maturity). Subsequent to initial fruit set and the period of active cell division (0–4 dpp) (Fu et al., 2010; Ando et al., 2012; Colle et al., 2017), fruit epidermal properties changed dramatically, especially in Gy14 (**Figures 1F–K**). Increases in epidermal cell height and width, cuticle thickness and intercalation between epidermal cells, and

lipid droplet number and size, generally showed a sigmoidal trend with fruit age. The greatest increases for most traits occurred between 4 and 12 dpp, coinciding with the period of exponential fruit growth (**Figure 1E**). Differences between Gy14 and CL9930 became apparent for most traits between 4 and 8 dpp and were largely stabilized by 16 dpp. Obvious intercalations and appearance of lipid droplet were observed sooner in Gy14, at 8 dpp, rather than 12 dpp in CL9930.

Fruit Epidermal QTLs

An $F_{7:8}$ Gy14 × CL9930 RIL population consisting of 110 lines was grown in the greenhouse in 2016 and evaluated for cucumber fruit epidermal traits as described above. Based on the observations of the developmental study, fruit were harvested at 16 dpp, after growth had stabilized and differences in fruit epidermal were readily observable. Phenotypic distributions and correlations among the traits are summarized in **Figure 2A**. Strong, positive correlations were observed among intercalation depth, epidermal cell height, and diameter and number of lipid droplets. Epidermal cell width was negatively correlated with epidermal cell height and number of lipid droplets.

Fourteen QTL were detected on six of the seven cucumber chromosomes (**Figure 3** and **Table 1**). On chromosome 1, a major QTL, *ECT1.1* (epidermal cell traits) was detected for cuticle thickness, epidermal cell height, intercalation depth, and diameter of lipid droplets that explained 18.4%, 38.1%, 44.1%, and 37.9% of the phenotypic variation for each trait, respectively. A single QTL was found on chromosome 2 for diameter of lipid droplets; chromosome 3 contained one QTL for epidermal cell height; and chromosome 4 had a single QTL for epidermal cell height, intercalation depth, epidermal cell width, and number of lipid droplets. In addition to the QTL for epidermal cell width found on chromosome 4, there also was a QTL detected on chromosome 5. Lastly, several QTL were found on chromosome 6 for intercalation depth, diameter of lipid droplets, and number of lipid droplets. In each case where there were multiple QTL for a single trait, the percent variation explained was greatest for the QTL on chromosome 1.

Marker-Assisted Screening and Fine Mapping of Chromosome 1

The strongest QTL were detected on chromosome 1, with LOD scores in the range of 6.1–28.7. Linkage analysis also supported the QTL on chromosome 1 (**Supplementary Table 5**). To narrow the region of interest, SSR markers were designed to flank the peak (at positions 14516668 and 18050191 CL9930 genome v3) with an additional marker in between (position 14783187). These markers were then used to screen an expanded $F_{7:8}$ RIL population (n = 375) to identify recombinant individuals in the region of interest. Of these, 87 lines were selected, including 29 identified as recombinant in the designated region on chromosome 1 (**Figure 4A**).

Selected lines, parents and reciprocal F_1s (Gy14 × CL9930 and CL9930 × Gy14) were grown in the field in 2018 and phenotyped at 20–22 dpp. With the exception of diameter of lipid droplets, the both F_1's showed intermediate phenotypes relative to the parents

(**Supplementary Figure 1**). Similar patterns of distribution and correlations among traits were observed for RILs as for the 2016 experiment in the greenhouse, with the exception of cuticle thickness, likely due to better imaging equipment in 2018 that allowed for more accurate determination of cuticle thickness (**Figure 2B**). In the 2018 experiment, cuticle thickness was strongly and positively correlated with intercalation depth, diameter of lipid droplets, and epidermal cell height (**Figure 2B**). The observed correlations among these four traits were consistent with their overlapping QTL positions on chromosome 1.

Combining phenotype data for the four traits with the SSR genotypic data, the region of interest on chromosome 1 was narrowed to approximately 3 Mb (14.78 Mb to 18.05). Recombinant individuals were then genotyped within this region with a set of seven SNP-based KASP™ markers spaced at approximately 0.5 Mb intervals. This narrowed the region of interest to 512 kb, from 16.76 to 17.28 Mb (**Figure 4B**). A subset of RILs recombinant in this region of chromosome 1 was also grown in the greenhouse in Spring 2019 to test expression of phenotype in different environments. Analysis of data from field Summer 2018 and greenhouse Spring 2019 showed very highly significant correlations (all P values <2.0E−06) between the two conditions for all four traits (cuticle thickness—r = 0.99; intercalation depth r = 0.96; diameter of lipid droplets r = 0.84; epidermal cell height r = 0.86) (**Supplementary Table 6**).

Candidate Genes Influencing Cucumber Fruit Epidermal Properties

The 0.51 Mb KASP marker-defined region on chromosome 1 contained 25 annotated genes (CL v3; http://cucurbitgenomics.org/). To refine the list of candidates we performed RNA-seq on fruit peels of the parental lines and further utilized existing expression data from our prior work comparing peels from 8 dpp and 16 dpp Gy14 fruit (Mansfeld et al., 2017) and expression data from Wei et al. (2016) comparing a wide variety of cucumber tissue types in CL9930 (accessed *via* http://cucurbitgenomics.org/, Gene expression project PRJNA 312872). As we were dealing with fruit epidermal related traits, and based on the developmental analyses showing that increase in cuticle related traits occurred most rapidly between 4 and 12 dpp, two criteria were used to filter the genes: (i) preferentially expressed in peel vs. flesh; and (ii) elevated expression at 8 dpp relative to 16 dpp. Of the 20 genes in this region showing expression in fruit, four had greater expression in peels than in flesh: *CsaV31g030090*, a putative heme oxygenase, associated with chlorophyll degradation; *CsaV31g030200*, a homolog of the cuticle-related transcription factor, *SHN1/WIN1*; *CsaV31g030210*, a gene with unknown function; and *CsaV31g030360*, a glucan endo-1,3-beta glucosidase (**Table 2**). These genes also showed greater expression in exocarp vs. mesocarp in our prior studies of pickling cucumber cv. Vlaspik, which also has a thick cuticle (Ando et al., 2015). Of the four genes, specificity to the peel was much stronger for *CsSHN1*, approximately 80-fold vs. 2-fold for *CsaV31g030090*, *CsaV31g030210*, and *CsaV31g030360*. Furthermore, *CsSHN1* was essentially exclusively expressed in fruit peel relative to other tissues and organs. In contrast, *CsaV31g030090* exhibited

FIGURE 2 | Correlation matrix containing scatterplots, distributions, and Pearson's correlation coefficients of the cucumber fruit epidermal traits for: **(A)** RIL population ($F_{7:8}$) of Gy14 × CL9930 (n = 110 lines, 16 dpp) grown in the greenhouse in Fall 2016. **(B)** Recombinant RIL lines (n = 87 lines, 20 dpp) grown in the field in 2018. Cuticle thickness (CT), Intercalation depth (ID), Diameter of lipid droplets (DLD), Epidermal cell height (ECH), Epidermal cell width (ECW), and Number of lipid droplets (NLD) in 450 μm linear region of epidermal cells. (*ns* not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. (Each value is the mean of three replicate fruit per RIL).

approximately 25-fold higher expression in flowers than fruit; *CsaV31g030210* exhibited 50–100-fold higher expression in roots, leaves, and flowers than in fruit; and *CsaV31g030360* was expressed comparably throughout the plant.

With regard to fruit development, *CsSHN1* was the only gene in the QTL1 region with significantly higher expression in peels of 8 dpp fruit than 16 dpp fruit (**Table 2**). Higher expression (P < 0.05 and 2-fold difference) of *CsSHN1* at 8 dpp than 16 dpp also was observed in the cultivars 'Vlaspik' and 'Poinsett 76' (**Figure 5A**). Examination of *CsSHN1* expression during cucumber fruit growth

from 0–20 dpp showed a sharp window of expression (8–12 dpp) during the period of exponential fruit growth, coinciding with peak cuticle and wax deposition (**Figure 5B**). Consistent with observed differences in chromosome 1-associated traits of cuticle thickness, intercalation depth, and diameter of lipid droplets, expression of *CsSHN1* was significantly higher in the two pickling cultigens, Gy14 and Vlaspik (Ando et al., 2015), than in CL9930. Peak expression was also somewhat delayed in CL9930, at 12 dpp vs. 8 dpp, corresponding with the relative timing for increase in cuticle thickness, intercalation depth, and diameter of lipid droplets.
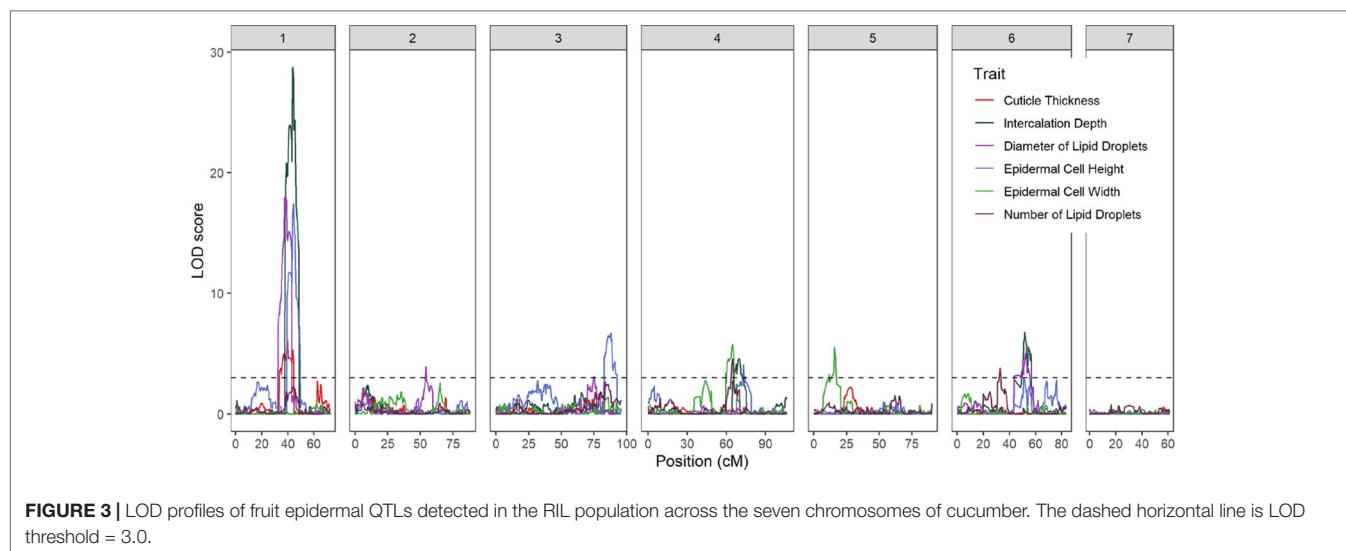
FIGURE 3 | LOD profiles of fruit epidermal QTLs detected in the RIL population across the seven chromosomes of cucumber. The dashed horizontal line is LOD threshold = 3.0.

TABLE 1 | Summary of fruit epidermal QTLs detected: Cuticle thickness (CT), Epidermal cell height (ECH), Intercalation depth (ID), Diameter of lipid droplets (DLD), Epidermal cell width (ECW), and Number of lipid droplets (NLD).

Trait	Chr	QTL Loci	Peak Pos (cM)	Nearest Marker	LOD Score	2.0 LOD Interval				Additive Effect (a)	Percent Variation (%) R^2
						Left Locus	Left cM	Right Locus	Right cM		
CT	1	qCT1.1	39.8	SNP.120657	6.1	SNP.10249	37.9	SNP.9737	43.1	0.34	18.4
ECH	1	qECH1.1	44.1	SNP.9649	17.4	SNP.9649	43.9	SNP.9341	44.9	−6.40	38.1
	3	qECH3.1	87.8	SNP.31473	6.7	SNP.32685	82.8	SNP.117081	89.1	3.60	11.5
	4	qECH4.1	73.0	SNP.128649	4.1	SNP.55545	71.6	SNP.146469	77.3	−2.69	6.7
ID	1	qID1.1	43.6	SNP.9649	28.7	SNP.9649	43.5	SNP.9649	44.1	−7.50	58.3
	4	qID4.1	69.8	SNP.55081	4.6	SNP.140945	67.9	SNP.128649	72.6	−2.07	4.5
	6	qID6.1	51.8	SNP.85409	6.8	SNP.84997	50.7	SNP.86689	53.1	−2.66	6.9
DLD	1	qDLB1.1	37.4	SNP.9941	18.0	SNP.9941	37.0	SNP.10249	37.9	−0.94	42.5
	2	qDLB2.1	54.2	SNP.143745	3.9	SNP.23461	52.8	SNP.131577	56.5	0.38	6.9
	6	qDLB6.1	52.2	SNP.175561	5.0	SNP.84997	50.8	SNP.128265	53.6	−0.44	8.8
ECW	4	qECW4.1	64.6	SNP.52733	5.8	SNP.55017	61.8	SNP.53205	65.7	2.23	14.8
	5	qECW5.1	15.7	SNP.76545	5.5	SNP.134473	15.2	SNP.77021	17.0	2.09	14.5
NLD	4	qNLB4.1	64.6	SNP.52733	4.6	SNP.122297	63.9	SNP.53205	65.5	−2.96	12.6
	6	qNLB6.1	32.8	SNP.151005	3.8	SNP.90761	30.8	SNP.89757	35.6	−3.53	10.3

Negative additive effect values (a) indicate that the allele is derived from parent Gy14. Positive additive effect values (a) indicate that the allele is derived from parent CL9930.

The predicted length of *CsSHN1* is 957 bp; transcript data (http://cucurbitgenomics.org/) support a single intron, consistent with other *SHINE* genes (Borisjuk et al., 2014). Comparison of the coding region plus 2 Kb upstream between Gy14 and CL9930 identified a SNP, within exon 2. The Vlaspik and Poinsett 76 sequences also shared the Gy14 sequence. A KASP marker was designed for the SNP at position 16961026 on chromosome 1 (CL9930 v. 3). The allele present (CL9930 vs. Gy14) at this position in the RILs completely co-segregated with phenotype. Marked allele effects were observed for the four fruit epidermal traits (**Figure 5C**).

The SNP at this position ('C' in Gy14 vs. 'G' in CL9930) results in a predicted amino acid change, from proline in Gy14 to arginine in CL9930, within a highly conserved region of the protein [domain CMV-1 as per Nakano et al. (2006)]. All of the other cucurbits for which there are draft genomes [*Citrullus lanatus, Cucumis melo, Cucurbita maxima, Cucurbita moschata, Cucurbita pepo, Cucurbita argyrosperma, Lagenaria siceraria*

(http://cucurbitgenomics.org/)] like Gy14, contain proline in this position (**Supplementary Table 5**). In addition, more than 30 divergent plant species with homologs identified by BLAST, also contain a proline residue at this position (**Supplementary Table 7**). Within cucumber germplasm, however, the CL9930 variant is quite common. Of 140 re-sequenced accessions with ≥10 reads at this position, 44 exhibited the CL9930 allele; another nine are heterozygous at this position (**Supplementary Table 4**).

DISCUSSION

Variation in Epidermal Properties of Cucumber Fruit During Development

Cucumber fruit sampled at incremental ages from anthesis through maturity were characterized for developmental changes and natural variation for epidermal traits, including epidermal
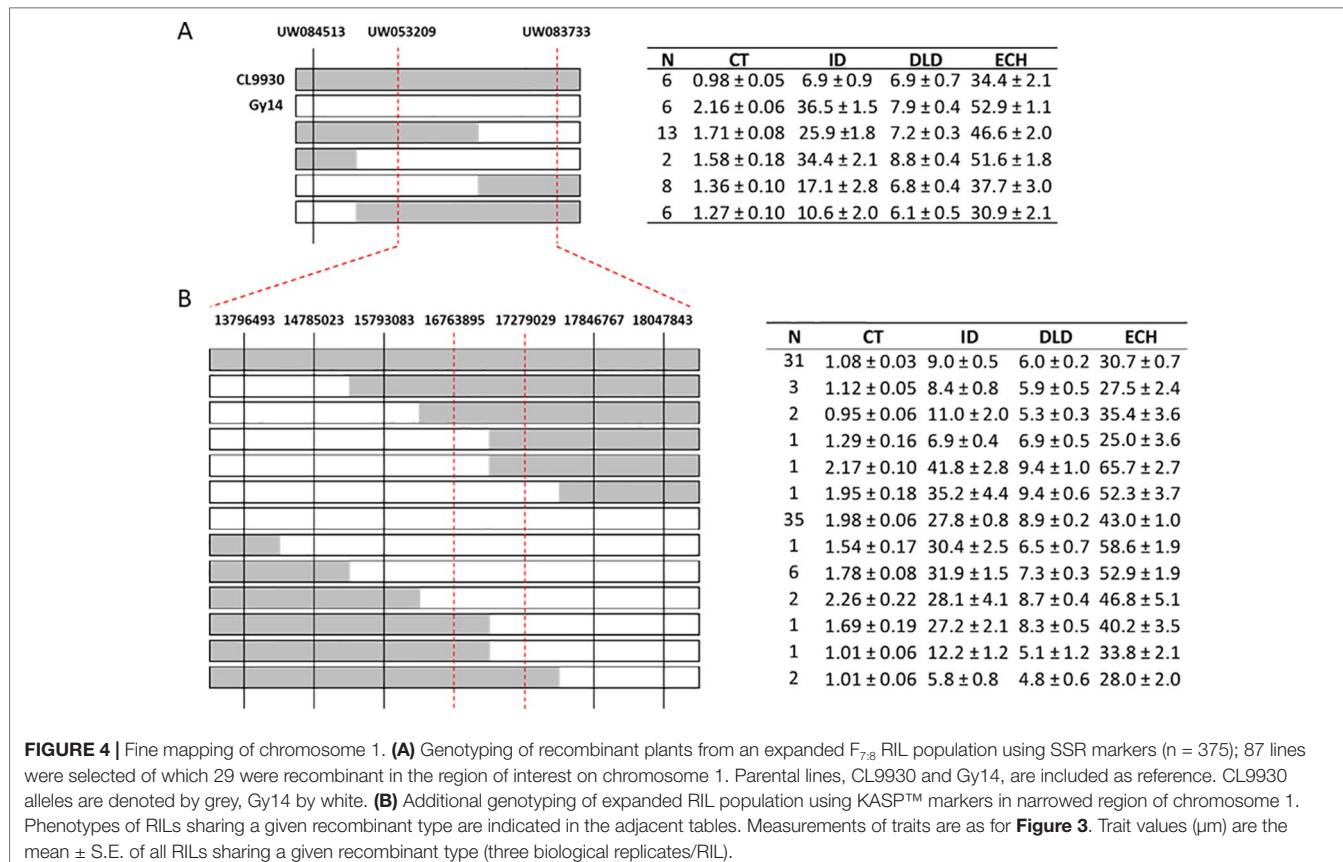
FIGURE 4 | Fine mapping of chromosome 1. **(A)** Genotyping of recombinant plants from an expanded $F_{7:8}$ RIL population using SSR markers (n = 375); 87 lines were selected of which 29 were recombinant in the region of interest on chromosome 1. Parental lines, CL9930 and Gy14, are included as reference. CL9930 alleles are denoted by grey, Gy14 by white. **(B)** Additional genotyping of expanded RIL population using KASP™ markers in narrowed region of chromosome 1. Phenotypes of RILs sharing a given recombinant type are indicated in the adjacent tables. Measurements of traits are as for **Figure 3**. Trait values (μm) are the mean ± S.E. of all RILs sharing a given recombinant type (three biological replicates/RIL).

cell shape, cuticle thickness, cuticular intercalations between epidermal cells, and the number and size of lipid droplets. Fruits from both Gy14 and CL9930 followed a characteristic sigmoidal pattern of growth, consistent with previous studies (Colle et al., 2017). The associated epidermal fruit traits also exhibited this pattern, with the greatest increase occurring at 4–12 days post pollination, coinciding with the period of exponential growth. The period of peak deposition of cuticle and wax during the period of maximal cucumber fruit growth is consistent with other systems where cuticle and wax deposition is developmentally programmed, often ceasing during early fruit development (Hen-Avivi et al., 2014; Martin and Rose, 2014; Lara et al., 2015; Trivedi et al., 2019). Beginning with the commencement of exponential growth and continuing throughout development, Gy14 had consistently larger values for cuticle thickness, cuticular intercalation, and number and size of lipid droplets. By 16 dpp, fruit size and differences in epidermal traits had largely plateaued; therefore, 16–20 dpp became the benchmark age for further epidermal work.

A striking observation was the presence of numerous large lipid droplets, typically 4–10 μm, in the epidermal cells. In plants, lipid droplets are thought to be formed in the endoplasmic reticulum and surrounded by a monolayer of phospholipids and structural membrane proteins (Chapman et al., 2012; Huang, 2018; Shimada et al., 2018). Lipid droplets in plants can vary quite widely in size, ranging from <1 to ~20 μm, depending on species and organ or tissue; larger ones are more

frequently found in oil rich fruit tissues (Goold et al., 2015). Much of what is known about the roles of lipid droplets comes from research involving seeds and leaves, but studies of fruits of avocado, olive, and oil palm suggest that the lipid droplets likely have varying functions for different tissue types (Pyc et al., 2017). Originally, lipid droplets were thought to have functions restricted to lipid storage, but recent findings have suggested that lipid droplets can be involved in more complex processes, such as lipid signaling and disease resistance (Chapman et al., 2012; Pyc et al., 2017). Lipid droplets also can sequester lipid-soluble compounds such as terpenoids that may contribute to protection against fungal or oomycete pathogens (e.g., Shimada et al., 2018; Sadre et al., 2019). Whether the cucumber fruit lipid droplets function in other capacities such as defense remains to be investigated.

Mapping of, and Relationships Among, Epidermal Fruit Traits

Phenotypic analysis of the Gy14 × CL9930 RIL populations was performed to ascertain genetic factors underlying the variation in epidermal traits. QTL for the six epidermal traits were detected on six of the seven cucumber chromosomes. Given the large LOD profiles and high correlation of traits mapping to chromosome 1, SSR markers designed to cover the peak QTL region on chromosomes 1 were used to screen an expanded Gy14 × CL9930 $F_{7:8}$ RIL population for recombinants in this region.

TABLE 2 | Annotated genes within fine-mapped region of QTL on cucumber chromosome 1 (16.764–17.279 Mb). Chromosomal positions and annotations are derived from Chinese Long v. 3.

Gene ID	Fruit expression (2 wks post-pollination)		Peel expression		Description		
	CL9930[a]		Gy14[b]			Arabidopsis Homolog	
	Flesh	Peel	8 dpp	16 dpp			e Value
CsaV31g030010	16	12	17.5	91.5	nitrate transporter	At2g26690	2.5 e–245
CsaV31g030020	12.2	9.7	11.2	15.1	similarity to DNA helicase	At1g65810	7.3 e–23
CsaV31g030030	0	0	0	0	Zn binding dehydrogenase family	At5g16990	3.5 e–23
CsaV31g030090	***13.9***	***21.7***	57.7	42.9	heme oxygenase	At2g26670	9.2 e–111
CsaV31g030110	8.5	2.2	2.4	4.7	metal transporter	At5g53550	3.2 e–286
CsaV31g030120	0	0	2.6	0.9	peroxidase superfamily	At2g41480	8.4 e–95
CsaV31g030140	0	0	0	0	TIR-NBS-LRR protein	At5g17680	7.2 e–157
CsaV31g030160	14.1	19.8	0	0	unknown	no match	
CsaV31g030170	0	0	0	0	peroxidase superfamily	At2g41480	4.3 e–117
CsaV31g030190	7.8	9.2	9.1	5.5	SPX domain containing	At2g26660	1.0 e–100
CsaV31g030200	***1.4***	***115.8***	***343.5***	***13.3***	SHINE1/WIN1	At1g15360	1.2 e–57
CsaV31g030210	***8.3***	***18.5***	7.2	11	unknown	At1g52565	2.1 e–12
CsaV31g030220	124.2	99.5	86.7	99.8	50S ribosomal L36	At5g20180	2.3e–30
CsaV31g030230	28.4	21.7	40.8	30.7	DUF 4050 family protein	At5g25360	2.5e–58
CsaV31g030250	0.1	0.5	0.5	0.1	EIN3 binding F-box protein-like	At5g25350	1.8e–84
CsaV31g030300	175.0	128.5	100.7	106.2	tetratricopeptide repeat protein	At5g20190	2.1e–25
CsaV31g030310	0.9	0.7	0.3	6.4	nucleotide transporter like	At1g02630	1.0e–133
CsaV31g030330	22.9	28.2	14.5	23.7	nuclear pore complex protein	At5g20200	7.8e–82
CsaV31g030340	0	0	0	0	axial regulator YABBY-like	At2g26580	2.4e–62
CsaV31g030350	8.7	8.1	13.8	7.6	methyl CpG binding domain	At3g15790	3.4e–29
CsaV31g030360	***6.8***	***13.1***	15.5	16.3	glucan endo-1-3 beta glucosidase	At2g26600	3.4e–29
CsaV31g030390	0.2	0	2.1	11.1	blue copper protein	At5g20230	4.7 e–16
CsaV31g030400	11.4	11.3	12.9	14.9	transducin family protein	At2g26610	0
CsaV31g030420	12.4	13.9	14.6	12.4	acetylgluosaminyl transferase family	At2g13290	6.8 e–143
CsaV31g030440	0	0	0	0	pentatricopeptide repeat containing	At4g02750	1.2 e–15

Predicted genes with no homology match and no expression in fruit were not included.

[a]*Expression data from Wei et al., 2016. Accessed via http://cucurbitgenomics.org/, Gene expression project PRJNA 312872.*

[b]*Expression data from Mansfeld et al., 2017. Available via http://cucurbitgenomics.org/, Gene expression project PRJNA 345040.*

Bold indicates higher expression in peel than flesh and higher expression at 8 dpp than 16 dpp.

Phenotyping of identified recombinants were narrowed to a region of ~3 Mb on chromosome 1. Fine mapping of the region of interest on chromosome 1 using KASP™ markers narrowed the region to an area of ~0.5 Mb. The very strong correlation between phenotype traits in the field and greenhouse indicate that, despite greatly differing environments, the measured traits on chromosome 1 are predominantly affected by genotype and developmental stage, rather than environment.

Strong, positive correlations were observed for cuticle thickness, intercalation depth, epidermal cell height, and diameter of lipid droplets along with a strong QTL on chromosome 1. While there was variation among RILS for relative intercalation depth (i.e., not all long cells had deep intercalations) the very strong correlation between epidermal cell height and intercalation depth argues that cell structure may be an important factor influencing cuticular intercalations in cucumber fruit. Diversity in intercalation patterns has been observed in a variety of species as illustrated by a 'grocery store survey' of numerous fruit types (Martin and Rose, 2014). The origin of such material remains unclear, possibly due to detachment from the epidermal cuticle and downward movement, or direct deposition to the anticlinal cell wall regions. It is also not clear whether diversity in intercalation patterns result from active regulation or from mechanical constraints of the cell structure (Martin and Rose, 2014). Consistent with a possible role of cell structure, in addition to the major QTL for intercalation depth and epidermal cell height on cucumber chromosome 1, intercalation depth also appears to share a QTL with epidermal cell height and cell width on chromosome 4.

Epidermal cell height and width showed a modest, negative association, and RIL phenotyping displayed a wide range of cell shapes beyond that of the flat and palisade orientations characteristic of CL9930 and Gy14, respectively. This variation is likely due to multiple factors controlling epidermal cell shape, including the *Pe* gene on chromosome 5. *Pe* has been localized to a 0.23 Mb region and exhibits tight, but not unbreakable, linkage to several fruit surface related traits such as dull (*D*), uniform fruit color (*u*) and tuberculate (*Tu*), suggesting a cluster of genes modulating cucumber exocarp characteristics (Yang et al., 2014a, Yang et al., 2014b, Yang et al., 2014c; Chen et al., 2016; Zhang et al., 2017; Yang et al., 2019). The QTL for epidermal cell width identified on chromosome 5 in this study is consistent with the location of *Pe*. Interestingly, the number of lipid droplets was not significantly related to size of lipid droplets suggesting that multiple factors regulate lipid droplet formation. Consistent with this observation, QTL for number of lipid body number and size were present on different chromosomes, 4 and 6 for number, and 1, 2 and 6, for size.
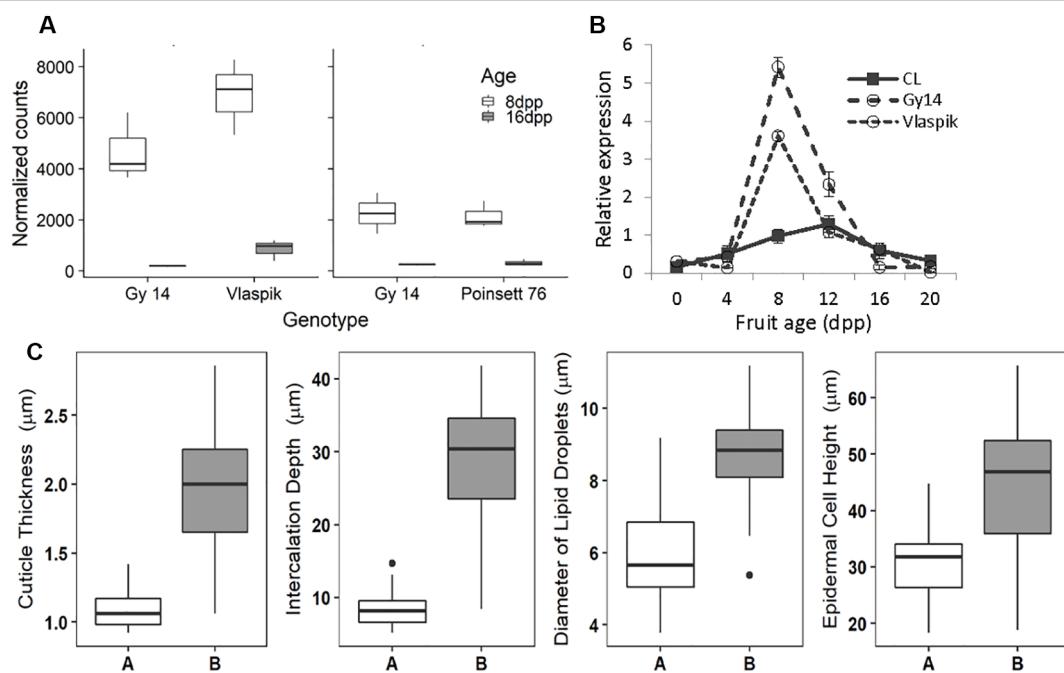
FIGURE 5 | Transcriptional analysis and allele effect of the CsSHINE1 (CsaV3_1g030200). **(A)** Comparison of expression of *CsSHN1* in fruit peel from cucumber fruit at 8 and 16 days post pollination (dpp). **(B)** Expression of *CsSHN1* in CL9930 and two pickling cultivars, Gy14 and Vlaspik, during cucumber fruit development, from anthesis (0 dpp) to 20 dpp. Each value is the mean of 3 biological replicates with 3 technical replicates/biological replicate ± S.E. **(C)** Allele effect of *CsSHN1* SNP at position 16961026 (in Chinese Long V3) for cuticle thickness, intercalation depth, diameter of lipid droplets and epidermal cell height as assessed from the expanded RIL population. 'A' refers to RILs with CL9930 allele (n = 25) and 'B' to RILs with the Gy14 allele (n = 35).

CsSHN1 is a Candidate Gene Influencing Cucumber Fruit Surface Properties

Mapping results and SSR and KASP marker assay refined the major QTL on chromosome 1 to a region containing 25 annotated genes. Expression profiles of these genes showing peak transcription coinciding with period of rapid fruit growth and deposition of cuticle, strongly preferential expression in fruit exocarp, and known function of SHINE transcription factors as regulators of cuticle and wax deposition (Yeats and Rose, 2013; Hen-Avivi et al., 2014; Trivedi et al., 2019), collectively implicate *CsSHN1* (Csa1g340430) as the primary candidate gene underlying the chromosome 1 QTL. *SHN* (*SHINE*) or *WIN* (*WAX INHIBITOR*) genes are members of the apetala2/ethylene-responsive element biniding protein (ap2/ere bp) transcription factor family originally named in Arabidopsis for their role in leaf appearance and the regulation of cuticle biosynthesis (Aharoni et al., 2004; Broun et al., 2004). *SHN* genes are primarily expressed in epidermal tissue in locations and periods of rapid growth, allowing for coverage and protection of the developing organ (Hen-Avivi et al., 2014; Trivedi et al., 2019). Expression of *CsSHN1* in cucumber fruit was consistent with this pattern, and mirrors the tissue specific and developmental regulation observed for *SlSHN3* in tomato fruit (Shi et al., 2013). Similar to *SlSHN3*, *CsSHN1* is nearly exclusively expressed in exocarp of immature fruit relative to other organs, tissues and ages.

Several studies have demonstrated that overexpression of *SHN* homologs increases wax deposition and cuticle thickness by modulating expression of cutin and wax biosynthesis genes, either directly or indirectly; conversely, down-regulation results in reduced cuticle and waxes (e.g., Aharoni et al., 2004; Broun et al., 2004; Kannangara et al., 2007; Shi et al., 2013). Variants for cuticle and wax deposition have primarily identified by mutant screens; however, more recent genomic, transcriptomic and metabolomic approaches have enabled the identification of natural variants (Cohen et al., 2017). While the majority of cuticle and wax variants identified to date include biosynthetic enzymes and lipid transporters [e.g., fatty acid omega hydroxylase (CYP861A, CYP86B1), BAHD acetyltransferase, beta-ketoacyl-CoA synthase, triterpene synthases, GDSL lipase] (Cohen et al., 2017), it has been suggested that regulatory genes are the most likely targets to achieve fine modulation (Petit et al., 2017). Naturally occurring variation for the naked caryopsis phenotype in barley, a trait causing loss of a sticky lipid substance secreted by the epidermis, was found to arise from mutation in a *SHN1* allele in barley (Taketa et al., 2008; Taketa et al., 2012) and genomic studies in apple have suggested that variations in the apple homolog of *SHN1* (*MdSHN3*), influence cuticle formation and russeting disorder in apple fruit (Lashbrooke et al., 2015b).

CsSHN1, like other SHINE and AP2/EREBP proteins includes the highly conserved ERF domain in the amino terminal portion of the protein. SHINE proteins are assigned to Group V of the

AP2/EREBP family (Nakano et al., 2006; Borisjuk et al., 2014). Group V includes a single intron and two conserved domains, CMV-1 and CMV-2, toward the middle and C-terminal portion of the protein, respectively. These features also occur in CsSHN1. The substitution of arginine for proline in CL9930 vs. Gy14 occurs within the conserved CMV-1 domain [also referred to as middle motif 'mm' (Aharoni et al., 2004)]. Mutation of a valine to aspartic acid mutation in this motif was shown to cause the naked caryopsis phenotype in barley, indicating functional significance of this domain (Taketa et al., 2008; Taketa et al., 2012). It remains to be determined whether the observed mutation in *CsSHN1* influences activity of the CsSHN1 transcription factor. The phenotypic differences observed between Gy14 and CL9930 may reflect protein activity and/or expression levels, as CL9930 also had reduced expression relative to Gy14. We did not observe sequence differences within the promoter (2 kb upstream of the coding region) or intron, suggesting that effects on transcript levels may result from other more distant elements or from relative RNA stability.

Despite conservation of the proline at this position among more than thirty species examined, including both dicots and monocots, the substitution was quite common among cucumber accessions (present in approximately a third of the re-sequenced lines). Cucumber is thought to have been first domesticated in South Asia and then subsequently moved both east toward China and west toward Europe, forming three major clades (Lv et al., 2012; Qi et al., 2013; Wang et al., 2018). Although there are exceptions, the CL9930 allele is predominantly found (70%) in East Asian accessions where it is widely present in cultivated East Asia cucumbers, but not landraces. This may suggest that this gene is under selection in the making of East Asia cucumbers (long, thin skin). The more frequent Gy14 allele is in present in many cultivars, landraces, semi-wild and wild cucumbers. Interestingly, though, it appears that the CL9930 allele also may be present at a relatively low frequency in the wild *C. sativus* var. *hardwickii*, as one of the 12 re-sequenced *hardwickii* accessions possessed this variant. This may reflect possible occurrence prior to domestication. Alternatively, as gene flow between cultivated cucumber and *hardwickii* populations occurs in natural populations (Bisht et al., 2004; Yang et al., 2012), this variation may have originated after domestication.

Several lines of evidence additionally suggest interplay between cuticle deposition and epidermal cell differentiation and development (e.g., Javelle et al., 2010; Nadakuduti et al., 2012; Yeats and Rose, 2013; Hen-Avivi et al., 2014; Fernandez-Moreno et al., 2017). This also has been observed for members of the *SHINE* family. For example, overexpression of Arabidopsis *SHN1* altered epidermal cell structure, including formation of elongated cells and reduced stomatal density and trichome number (Aharoni et al., 2004) and tomato *SlSHN3* influences cell shape, either directly, or indirectly by influencing expression of other cell downstream patterning genes such as *SlMIXTA* (Shi et al., 2013; Lashbrooke et al., 2015a). Down regulation of *SlSHN3* results in reduced cuticle deposition and flattened fruit epidermal cells. The connection between *SHINE*-family member genes and cell shape may also contribute to the observed QTL for cucumber epidermal cell height at this location.

CONCLUSIONS

Cucumber fruit epidermis exhibits dynamic developmental changes during fruit growth including changes in cell size and shape, deposition of cuticle, and appearance of lipid droplets. There is also natural variation for these traits as manifest in differing cucumber market classes, and observed for the Chinese fresh market cucumber, CL9930, relative to the American pickling cucumber, Gy14. Genetic analyses indicated several QTL, including a major QTL on chromosome 1, QTL *ECT1.1*, influencing cuticle thickness and depth of intercalation between epidermal cells, diameter of lipid droplets and epidermal cell height. Additional QTL of lesser impact were present on chromosomes 3, 4, 5 and 6. Fine mapping of the four traits associated with QTL *ECT1.1* narrowed the region to 0.5 Mb. Transcriptomic analysis based on tissue-specific and developmentally-regulated expression of fruit epidermal traits of genes in this region along with and observed allelic effects, identified a primary candidate gene—a homolog of *SHINE1,* which in other systems has been shown to influence both cuticle deposition and epidermal cell shape. The *CsSHN1* sequence in CL9930 includes a single base difference causing an amino acid change (proline to arginine) in the highly conserved CMV-1 domain when compared to that in Gy14. This single base change, which occurred frequently in East Asian cucumber accessions may contribute to natural variation for cucumber epidermal properties. As epidermal properties, including wax deposition, influence both consumer preferences and longevity in the market chain, allelic variation in *CsSHN1* may provide a valuable target for breeders developing varieties to meet desired fruit quality characteristics, such as fruit with shinier appearance (reduced wax) or extended shelf life due to reduced water loss (increased wax).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI Sequence Read Archive (SRA) under project accession number PRJNA558838.

AUTHORS CONTRIBUTIONS

SR-C, RG and CB conceived of the project. SR-C and RG wrote the paper. SR-C performed the developmental analyses, phenotyping of the RILs, QTL mapping, KASP analyses and fine mapping. MC performed the developmental analysis of CsSHN1 expression. BM performed the RNA-Seq experiment, differential gene expression analysis and SNP calling of parental lines. YWa and YWe performed the SSR analysis and identification of recombinants. LG and ZF analyzed the re-sequenced cucumber accessions for alleles of CsSHN1.

FUNDING

REFERENCES

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. (2009). Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. *PloS Genet.* 5, e1000777. doi: 10.1371/journal.pgen.1000777

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., and Pereiraa, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in Arabidopsis. *Plant Cell* 16, 2463–2480. doi: 10.1105/tpc.104.022897

Ando, K., Carr, K. M., and Grumet, R. (2012). Transcriptome analysis of early cucumber fruit growth identifies distinct gene modules associated with phases of development. *BMC Genomics* 13, 518. doi: 10.1186/1471-2164-13-518

Ando, K., Carr, K. M., Colle, M., Mansfeld, B. N., and Grumet, R. (2015). Exocarp properties and transcriptomic analysis of cucumber (*Cucumis sativus*) fruit expressing resistance to *Phytophthora capsici*. *PloS One* 10, e0142133. doi: 10.1371/journal.pone.0142133

Ando, K., Grumet, R. 2010. Transcriptional profiling of rapidly growing cucumber fruit by 454-pyrosequencing analysis. *J. Amer. Soc. Hortic. Sci.* 135, 291–302. doi: 10.21273/JASHS.135.4.291

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc B.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bisht, I. S., Bhat, K. V., Tanwar, S. P. S., Bhandari, D. C., Joshi, K., and Sharma, A. K. (2004). Distribution and genetic diversity of *Cucumis sativus* var. *hardwickii* (Royle) Alef in India. *J. Hortic. Sci. Biotech.* 79, 783–791. doi: 10.1080/14620316.2004.11511843

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

Borisjuk, N., Hrmova, M., and Lopato, S. (2014). Transcriptional regulation of cuticle biosynthesis. *Biotechnol. Adv.* 32, 526–540. doi: 10.1016/j.biotechadv.2014.01.005

Broun, P., Poindexter, P., Osborne, E., Jiang, C.-Z., and Riechmann, J. L. (2004). WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4706–4711. doi: 10.1073/pnas.0305574101

Buda, G. J., Isaacson, T., Matas, A. J., Paolillo, D. J., and Rose, J. K. C. (2009). Three-dimensional imaging of plant cuticle architecture using confocal scanning laser microscopy. *Plant J.* 60, 378–385. doi: 10.1111/j.1365-313X.2009.03960.x

Chapman, K. D., Dyer, J. M., and Mullen, R. T. (2012). Biogenesis and functions of lipid droplets in plants. *J. Lipid Res.* 53, 215–226. doi: 10.1194/jlr.R021436

Chen, C. H., Yin, S., Liu, X. W., Liu, B., Yang, S., Xue, S. D., et al. (2016). The WD-repeat protein CsTTG1 regulates fruit wart formation through interaction with the Homeodomain-Leucine zipper I protein Mict. *Plant Physiol.* 171, 1156–1168. doi: 10.1104/pp.16.00112

Cohen, H., Szymanski, J., and Aharoni, A. (2017). Assimilation of 'omics' strategies to study the cuticle layer and suberin lamellae in plants. *J. Exp. Bot.* 68, 5389–5400. doi: 10.1093/jxb/erx348

Colle, M., Weng, Y., Kang, Y., Ophir, R., Sherman, A., and Grumet, R. (2017). Variation in cucumber (*Cucumis sativus* L.) fruit size and shape results from multiple components acting pre-anthesis and post-pollination. *Planta* 246, 641–658. doi: 10.1007/s00425-017-2721-9

Colle, M. (2015). *Cucumber (Cucumis sativus L.) fruit development: factors influencing fruit size, shape and resistance to Phytophthora capsisi.* East Lansing MI: PhD Dissertation, Michigan State University.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806

Fernandez-Moreno, J. P., Levy-Samoha, D., Malitsky, S., Monforte, A. J., Orzaiz, D., Aharoni, A., et al. (2017). Uncovering tomato quantitative trait loci and candidate genes for fruit cuticular lipid composition using the *Solanum pennellii* introgression line population. *J. Exp. Bot.* 68, 2703–2716. doi: 10.1093/jxb/erx134

Fu, F. Q., Mao, W. H., Shi, K., Zhou, Y. H., and Yu, J. Q. (2010). Spatio-temporal changes in cell division, endoreduplication and expression of cell cycle-related genes in pollinated and plant growth substances-treated ovaries of cucumber. *Plant Biol.* 12, 98–107. doi: 10.1111/j.1438-8677.2009.00203

Gao, M. L., Hu, L. L., Li, Y. H., 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

Gimenez, E., Dominguez, E., Pineda, B., Heredia, A., Moreno, V., Lozano, R., et al. (2015). Transcriptional activity of the MADS box *ARLEQUIN/TOMATO AGAMOUS-LIKE1* gene is required for cuticle development of tomato fruit. *Plant Physiol.* 168, 1036–1048. doi: 10.1104/pp.15.00469

Goold, H., Beisson, F., Peltier, G., and Li-Beisson, Y. (2015). Microalgal lipid droplets: composition, diversity, biogenesis and functions. *Plant Cell Rep.* 34, 545–555. doi: 10.1007/s00299-014-1711-7

Hen-Avivi, S., Lashbrooke, J., Costa, F., and Aharoni, A. (2014). Scratching the surface: genetic regulation of cuticle assembly in fleshy fruit. *J. Exp. Bot.* 65, 4653–4664. doi: 10.1093/jxb/eru225

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, 1275–1281. doi: 10.1038/ng.475

Huang, A. H. C. (2018). Plant lipid droplets and their associated proteins: potential for rapid advances. *Plant Physiol.* 176, 1894–1918. doi: doi.org/10.1104/pp.17.01677

Ingram, G., and Nawrath, C. (2017). The roles of the cuticle in plant development: organ adhesions and beyond. *J. Exp. Bot.* 68, 5307–5321. doi: 10.1093/jxb/erx313

Isaacson, T., Kosma, D. K., Matas, A. J., Buda, G. J., He, Y., Yu, B., et al. (2009). Cutin deficiency in the tomato fruit cuticle consistently affects resistance to microbial infection and biomechanical properties, but not transpirational water loss. *Plant J.* 60, 363–377. doi: 10.1111/j.1365-313X.2009.03969.x

Javelle, M., Vernoud, V., Rogowsky, P. M., and Ingram, G. C. (2010). Epidermis: the formation and functions of a fundamental plant tissue. *New Phytol.* 189, 17–39. doi: 10.1111/j.1469-8137.2010.03514

Kannangara, R., Branigan, C., Liu, Y., Penfield, T., Rao, V., Mouille, G., et al. (2007). The transcription factor WIN1/SHN1 regulates cutin biosynthesis in *Arabidopsis thaliana*. *Plant Cell.* 19, 1278–1294. doi: 10.1105/tpc.106.047076

Kunst, L., and Samuels, A. L. (2003). Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Res.* 42, 51–80. doi: 10.1016/s0163-7827(02)00045-0

Lara, I., Belge, B., and Goulao, L. F. (2014). The fruit cuticle as a modulator of postharvest quality. *Postharv. Biol. Tech.* 87, 103–112. doi: 10.1016/j.postharvbio.2013.08.012

Lara, I., Belge, B., Goulao, L. F. (2015). A focus on biosynthesis and composition of cuticle in fruits. *JAFC* 63, 4005–4019. doi: 10.1021/acs.jafc.5b00013

Lashbrooke, J., Adato, A., Lotan, O., Alkan, N., Tsimbalist, T., Rechav, K., et al. (2015a). The tomato MIXTA-Like transcription factor coordinates fruit epidermis conical cell development and cuticular lipid biosynthesis and assembly. *Plant Physiol.* 169, 2552–2571. doi: 10.1104/pp.15.01145

Lashbrooke, J., Aharoni, A., and Costa, F. (2015b). Genome investigation suggests MdSHN3, an APETALA2-domain transcription factor gene, to be a positive regulator of apple fruit cuticle formation and an inhibitor of russet development. *J. Exp. Bot.* 66, 6579–6589. doi: 10.1093/jxb/erv366

Lauter, N., Kampani, A., Carlson, S., Goebel, M., and Moose, S. P. (2005). microRNA172 down-regulates *glossy15* to promote vegetative phase change in maize. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9412–9417. doi: 10.1073/pnas.0503927102

Li, Z., Zhang, Z. H., Yan, P. C., Huang, S. W., Fei, Z. J., 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

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

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, 550–571. doi: 10.1186/S13059-014-0550-8

Lv, J., Qi, J., Shi, Q., Shen, D., Zhang, S., Shao, G., et al. (2012). Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). *PLoS One* 7, e46919. doi: 10.1371/journal.pone.0046919

Mansfeld, B. N., Colle, M., Kang, Y., Jones, A. D., and Grumet, R. (2017). Transcriptomic and metabolomic analyses of cucumber fruit peels reveal a developmental increase in terpenoid glycosides associated with age-related resistance to *Phytophthora capsici*. *Hortic. Res.* 4, 17022. doi: 10.1038/hortres.2017.22

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), e1005944. doi: 10.1371/journal.pcbi.1005944

Martin, L. B. B., and Rose, J. K. C. (2014). There's more than one way to skin a fruit: formation and functions of fruit cuticles. *J. Exp. Bot.* 65, 4639–4651. doi: 10.1093/jxb/eru301

Matas, A. J., Agusti, J., Tadeo, F. R., Talon, M., and Rose, J. K. C. (2010). Tissue-specific transcriptome profiling of the citrus fruit epidermis and subepidermis using laser capture microdissection. *J. Exp. Bot.* 61, 3321–3330. doi: 10.1093/jxb/erq153

Matas, A. J., Yeats, T. H., Buda, G. J., Zheng, Y., Chatterjee, S., Tohge, T., et al. (2011). Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *Plant Cell* 23, 3983–3910. doi: 10.1105/tpc.111.091173

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

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldber, L., Lotan, O., Yativ, M., et al. (2008). Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol.* 147, 823–851. doi: 10.1104/pp.108.116004

Moose, S. P., and Sisco, P. H. (1996). *Glossy15*, an APETALA2-like gene from maize that regulates leaf epidermal cell identity. *Genes Dev.* 10, 3018–3027. doi: 10.1101/gad.10.23.3018

Nadakuduti, S. S., Pollard, M., Kosma, D. K., Allen, C.Jr., Ohlrogge, J. B., and Barry, C. S. (2012). Pleiotropic phenotypes of the sticky peel mutant provide new insight into the role of CUTIN DEFICIENT2 in epidermal cell function in tomato. *Plant Physiol.* 159, 945–960. doi: 10.1104/pp.112.198374

Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol.* 140, 411–432. doi: 10.1104/pp.105.073783

Oshima, Y., Shikata, M., Koyama, T., Ohtsubo, N., Mitsuda, N., and Ohme-Takagi, M. (2013). MIXTA-like transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle development in Arabidopsis and *Torenia fournieri*. *Plant Cell.* 25, 1609–1624. doi: 10.1105/tpc.113.110783

Patel, C., and Panigrahi, J. (2019). Starch glucose coating-induced postharvest shelf-life extension of cucumber. *Food Chem.* 288, 208–214. doi: 10.1016/j.foodchem.2019.02.123

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

Petit, J., Bres, C., Just, D., Garcia, V., Mauxion, J. P., Marion, D., et al. (2014). Analyses of tomato fruit brightness mutants uncover both cutin-deficient and cutin-abundant mutants and a new hypomorphic allele of GDSL lipase. *Plant Physiol.* 164, 888–906. doi: 10.1104/pp.113.232645

Petit, J., Bres, C., Mauxion, J.-P., Bakan, B., and Rothan, C. (2017). Breeding for cuticle-associated traits in crop species: traits, target, and strategies. *J. Exp. Bot.* 68, 5369–5387. doi: 10.1093/jxb/erx341

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

Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45, 1510–1515. doi: 10.1038/ng.2801

Rubenstein, M., Katzenellenbogen, M., Exhed, R., Rozen, A., Katzir, N., Colle, M., et al. (2015). Ultrahigh-density linkage map for cultivated cucumber (*Cucumis sativus* L.) using a single-nucleotide polymorphism genotyping array. *PLoS One* 10, e0124101. doi: 10.1371/journal.pone.0124101

Sadre, R., Kuo, P., Chen, J., Yang, Y., Banerjee, A., Benning, C., et al. (2019). Cytosolic lipid droplets as engineered organelles for production and accumulation of terpenoid biomaterials in leaves. *Nat. Commun.* 10, 853. doi: 10.1038/s41467-019-08515-4

Shi, J. X., Adato, A., Alkan, N., He, Y., Lashbrooke, J., Matas, A. J., et al. (2013). The tomato *SlSHINE3* transcription factor regulates fruit cuticle formation and epidermal patterning. *New Phytol.* 197, 468–480. doi: 10.1111/nph.12032

Shimada, T. L., Hayashi, M., and Hara-Nishimura, I. (2018). Membrane dynamics and multiple functions of oil bodies in seeds and leaves. *Plant Physiol.* 176, 199–207. doi: 10.1104/pp.17.01522

Soneson, C., Love, M. I., Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4, 1521. doi: 10.12688/f1000research.7563.2

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kadeda, K., et al. (2008). Barley grain with adhering hulls in controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proc. Nat. Acad. Sci. U.S.A.* 105, 4062–4067. doi: 10.1073./pnas.0711034105

Taketa, S., Yuo, T., Yamashita, Y., Ozeki, M., Haruyama, N., Hidekazu, M., et al. (2012). "Molecular mechanisms for covered vs. naked caryopsis in barley. *Adv. Barley Sci.*," in *Advance in Barley Sciences*. Eds. Zhang, G., Li, C., and Liu, X. (Dordrecht: Springer). doi: 10.1007/978-94-007-4682-4_39

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

Trivedi, P., Nguyen, N., Hykkerud, A. L., Haggman, H., Martinussen, I., Jaakola, L., et al. (2019). Development and environmental regulation of cuticular wax biosynthesis in fleshy fruits. *Front. Plant Sci.* 10, 431. doi: 10.3389/fpls.2019.00431

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). "From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline," in *Curr. Protoc. Bioinforma* (Hoboken, NJ, USA: John Wiley & Sons, Inc.) 11.10.1–11.10.33.

Wang, S., Basten, C. J., and Zeng, Z.-B. (2012). *Windows QTL Cartographer 2.5*. Raleigh, NC: Department of Statistics, North Carolina State University. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Wang, W. Liu, X., Gai, X., Ren, J., Liu, X., Cai, Y., Wang, Q., and Ren, R. (2015a). Cucumis sativus L. WAX2 plays a pivotal role in wax biosynthesis, influencing pollen fertility and plant biotic and abiotic stress responses. *Plant Cell Physiol.* 56, 1339–1354. doi: 10.1093/pcp/pcv052

Wang, W. J., Zhang, Y., Xu, C., Ren, J. J., Liu, X. F., Black, K., et al. (2015b). Cucumber *ECERIFERUM1* (*CsCER1*), which influences the cuticle properties and drought tolerance of cucumber, plays a key role in VLC alkanes biosynthesis. *Plant Molec. Biol.* 87, 219–233. doi: 10.1007/s11103-014-0271-0

Wang, X., Bao, K., Reddy, U. K., Bai, Y., Hammar, S. A., Jiao, C., et al. (2018). The USDA cucumber (*Cucumis sativus* L.) collection: genetic diversity, population structure, genome-wide association studies and core collection development. *Hortic. Res.* 5, 64. doi: 10.1038/s41438-018-0080-8

Wei, G., Tian, P., Zhang, F., Qin, Q., Miao, H., Chen, Q., et al. (2016). Integrative analysis of nontargeted volatile profiling and transcriptome data provide molecular insight into VOC diversity in cucumber plants (*Cucumis sativus*). *Plant Physiol.* 172, 603–618. doi: 10.1104/pp.16.01051

Weng, Y., Colle, M., Wang, Y. H., Yang, L. M., Rubinstein, M., Sherman, A., et al. (2015). QTL mapping in multiple populations and development stages reveals dynamic quantitative trait loci for fruit size in cucumbers of different market classes. *Theor. Appl. Genet.* 128, 1747–1763. doi: 10.1007/s00122-015-2544-7

Wiersma, A. T., Pulman, J. A., Brown, L. K., Cowger, C., and Olson, E. L. (2017). Identification of Pm58 from *Aegilops tauschii. Theor. Appl. Genet.* 130, 1123–1133. doi: 10.1007/s00122-017-2874-8

Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313X.2012.05017

Yang, X. Q., Yue, L., Zhang, W. W., He, H. L., Pan, J. S., and Cai, R. (2014a). Fine mapping of the uniform immature fruit color gene u in cucumber (*Cucumis sativus* L.). *Euphytica.* 196, 341–348. doi: 10.1007/s10681-013-1037-5

Yang, X., Zhang, W., Li, Y., He, H., Bie, B., Ren, G., et al. (2014b). mapping of the dull fruit skin gene D in cucumber (*Cucumis sativus* L). *Mol. Breed.* 33, 15–22. doi: 10.1007/s11032-013-9927-8

Yang, X., Zhang, W., He, H., Nie, J., Bie, B., Zhao, J., et al. (2014c). Tuberculate fruit gene *Tu* encodes a C2H2 zinc finger protein that is required for the warty fruit phenotype in cucumber (*Cucumis sativus* L.). *Plant J.* 78, 1034–1046. doi: 10.1111/tpj.12531

Yang, S., Wen, C., Liu, B., Cai, Y., Xue, S., Bartholomew, E. Z., et al. (2019). A CsTu-Ts1 regulatory module promotes fruit tubercule formation in cucumber. *Plant Biotech. J.* 17, 289–301. doi: 10.1111/pbi.12977

Yeats, T. H., and Rose, J. K. (2013). The formation and function of plant cuticles. *Plant Physiol.* 163, 5–20. doi: 10.1104/pp.113.222737

Yeats, T. H., Martin, L. B., Viart, H. M., Isaacson, T., He, Y., Zhao, L., et al. (2012). The identification of cutin synthase: formation of the plant polyester cutin. *Nat. Chem. Biol.* 8, 609–611. doi: 10.1038/nchembio.960

Zhang, S., Miao, H., Song, Z. C., Liu, P. N., Wang, Y., Wehner, T. C., et al. (2017). Molecular mapping and candidate gene analysis for fruit epidermal structure in cucumber. *Plant Breed.* 136, 767–774. doi: 10.1111/pbr.12510

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. *Nucl. Acids Res.* 47, D1128–D1136. doi: 10.1093/nar/gky944

Transcription Factor *CsWIN1* Regulates Pericarp Wax Biosynthesis in Cucumber Grafted on Pumpkin

Jian Zhang[1,2], Jingjing Yang[1,2], Yang Yang[1,2], Jiang Luo[1,2], Xuyang Zheng[1,2], Changlong Wen[1,2]* and Yong Xu[1,2]*

[1] Beijing Vegetable Research Center (BVRC), Beijing Academy of Agricultural and Forestry Sciences, National Engineering Research Center for Vegetables, Beijing, China, [2] Beijing Key Laboratory of Vegetable Germplasms Improvement, Beijing, China

Pericarp wax of cucumber is an important economic trait, determining sales and marketing. Grafting of cucumber onto pumpkin rootstock (*Cucurbita moschata*) is an effective way to produce glossy cucumber fruits. However, the molecular regulation mechanism of this phenomenon remains largely unknown. In the present study, transcriptome analyses, genome-wide DNA methylation sequencing, and wax metabolite analysis were performed on the pericarp of self-rooted versus grafted cucumber. We identified the AP2/ERF-type transcription factor *CsWIN1* as methylated and significantly upregulated in grafted cucumber compared to self-rooted cucumber. The increased expression of *CsWIN1* was also positively correlated with several key wax biosynthesis genes, including *CsCER1*, *CsCER1-1*, *CsCER4*, *CsKCS1*, and the wax transporter gene *CsABC*. The transcriptome expression level of these genes was validated through qRT-PCR profiles. Furthermore, wax metabolite analysis showed that more wax ester (C20 fatty acid composition), but fewer alkanes (C29 and C31) were deposited in grafted cucumber pericarp. The higher expression of *CsWIN1* and wax biosynthesis genes was reflected in the glossier appearance of grafted pericarp, possibly the result of higher wax ester content and higher integration of small trichomes in the pericarp. This study demonstrates that grafting can affect the content and composition of pericarp wax in cucumber grafted on pumpkin, and a unique regulation model of *CsWIN1* for wax biosynthesis may exist in cucumber.

Keywords: pericarp wax, transcriptome, methylation, *CsWIN1*, grafted cucumber

INTRODUCTION

Cucumber is one of the major vegetables cultivated worldwide (Huang et al., 2009). China is the largest producer and consumer of cucumber with about 77.4% of the total cucumber yield in the world (Year 2017, www.fao.org/faostat/en/).

Glossy cucumber is a dominant player in the Chinese market. The pericarp appearance of cucumber is an important economical factor, appealing to customers and plays an important role in sales and marketing of the fruit worldwide. The pericarp wax is mainly composed of wax and cutin derived from the cuticle and is a critical trait affecting the appearance and quality of cucumber (Wang et al., 2015b; Li et al., 2019). The waxy cucumber has a natural white powder-like substance on the pericarp surface, which is less desirable than cucumber varieties with a glossy pericarp.

Tian et al., (2015) identified five QTL (*WP1.1*, *WP3.1*, *WP5.1*, *WP6.1* and *WP6.2*) promoting pericarp wax accumulation and two QTL (*WP5.1* and *WP6.2*) with moderate effect. Other researchers revealed that the pericarp wax in cucumber is affected by both potential biosynthesis genes (*CsCER1*, *CsWAX2*) and environmental factors, such as light intensity, temperature, moisture, and grafting (Liu et al., 2014; Wang et al., 2015b; Wang et al., 2015c). However, there has been little research focusing on the genetic mapping of pericarp wax biosynthesis genes, which are critical to the discovery of the pericarp wax biosynthesis pathway in cucumber. Grafting onto pumpkin rootstock (*Cucurbita moschata*) proved to be an effective way to brighten the pericarp of cucumber (Samuels et al., 1993; Buschhaus et al., 2015; Shen et al., 2015). Still, the molecular mechanisms in regulating pericarp wax biosynthesis in cucumber grafted on pumpkin is unknown. There is a need to elucidate the biosynthesis pathway of pericarp wax in cucumber and identify the gene loci that confer glossy appearance in cucumber.

The cuticle layer covers the epidermal cells of leaves and other plant parts. It protects the plant from various environmental stresses, including fungal pathogens, UV radiation, water loss and non-stomatal transpiration, and dust deposits (Kunst and Samuels, 2009; Buschhaus et al., 2015; Blanc et al., 2018).The chemical components of cuticular wax has been well studied in *Arabidopsis* and other crops. Several functional genes involved in wax biosynthesis (*CER1*, *WAX2*, and *FAC4*) and transportation (ABC/LTP transporters) were identified (Broun et al., 2004; Suh et al., 2005; Bernard et al., 2012; Park et al., 2016). Cuticular wax was shown to be biosynthesized from C16 and C18-CoAsin plastids and then elongated into a long chain fatty acids (VLCFAs, chain length >20 carbons) in the endoplasmic reticulum membrane (Suh et al., 2005; Kunst and Samuels, 2009). Cuticular wax covers the outer surface of the cutin layer and presents in the form of crystals, a complex mixture that contains various primary and secondary alcohols, alkenes, ketones, and esters derived from VLCFAs through alkane- and alcohol-forming pathways (Lee et al., 2009; Gan et al., 2016). However, the chemical composition and content of cuticular wax are different among species and even tissues (Kunst and Samuels, 2009). Moreover, several studies have found that cuticular wax biosynthesis is affected by environmental factors, such as water deficit, lower temperature, and high levels of UV light (Kunst and Samuels, 2009; Lee and Mi, 2015). AP2-ERF (APETALA2/ethylene) transcription factors (TFs) were involved in the regulation of ET-response genes and had proved to possess a variety of functions in plants (Gutterson and Reuber, 2004; Feng et al., 2005; Li et al., 2011; Huang et al., 2016). Among these, *WIN1* was first identified in a mutant in *Arabidopsis thaliana*; the overexpression of *WIN1* caused a glossy phenotype, indicating that *WIN1* is a regulator of wax biosynthesis, and it was reported to activate the expression of wax biosynthesis genes such as *CER1*, *CER2*, and *KCS1*(Broun et al., 2004). The other AP2/ERF gene *DEWAX* was found to negatively regulate cuticular wax biosynthesis in *A. thaliana* (Go et al., 2014). Furthermore, a positive regulator of wax biosynthesis "*WRI4*", a member of the ERF family was identified in *Arabidopsis* stems (Park et al., 2016). There is a need to determine the role of each of the AP2/ERF transcription factors in regulating wax biosynthesis in cucumber.

Glossy cucumber is a dominant player in the Chinese market. Grafting is used for defending cucumber plants against soil-borne diseases. At the same time, it is an important method for production of glossy cucumbers. In this study, the cucumber variety "Jingyan 118" was grafted onto the pumpkin rootstock variety "Jingxinzhen 6," an elite line used in grafting to brighten the scion cucumber pericarp. We conducted associated transcriptome and genome-wide methylation analyses in conjunction with changes in the cucumber pericarp wax in response to grafting. We determined that the *CsWIN1* gene of cucumber (homologous to *WIN1* in *Arabidopsis*) is methylated and upregulated and at the same time transcriptionally regulates several wax biosynthesis genes, including *CsCER1* and *CsCER4* in the pericarp of grafted cucumber.

MATERIALS AND METHODS

Plant Materials and Grafting Experiment

The experiment was conducted in the Beijing Vegetable Research Center (BVRC) from 26 March to 10 July, 2017. Seeds of the cucumber variety "Jingyan 118" with high pericarp wax and a pumpkin rootstock "Jingxinzhen 6" were sown in a standard potting mix (peat: sand: pumice, 1:1:1, V/V/V). The cuttage grafting system was applied when the scions were growing at the one true leaf and the cotyledon of the rootstock was in the expansion stage. To enhance the survival rate, grafted seedlings were kept in the shade (24–28°C, 80–90% RH) for 3 days. Two weeks later, self-rooted and grafted seedlings were transplanted into soil in a greenhouse. Fertilization and cultivation management methods were as commonly recommended in cucumber production. The pericarp of self-rooted and grafted cucumber at commodity maturity were extracted for the following experiments below.

Pericarp Wax Observation and Chemical Component Analysis

Using a sharp thin blade, a 1 cm² pericarp was carefully cut off from cucumber fruits at marketable mature stage. Images of the cuticular wax crystals were visualized at 200× magnification using a scanning electron microscope (Semagn et al., 2014) (S4700, Hitachi, Japan). Cellular morphology under the microscope was also observed using cryosection techniques. Long alkanes analysis of wax by gas chromatography was performed, following a method described by Park (Park et al., 2016). Wax esters containing saturated and unsaturated fatty acids from five biological replicates were detected by using specific multiple-reaction monitoring (MRM) scanning (Lam et al., 2013).

RNA Isolation and Library Construction

Total RNA of the pericarp from self-rooted, grafted, and failed grafted cucumber (cucumber scion which developed roots connected with the stock pumpkin and/or raised new roots into the soil) with three biological replicates were extracted using a DNeasy Kit and miRNeasy Kit, respectively (QIAGEN, USA). The concentration and quality of DNA and RNA were

evaluated by a NanoDrop 2000C Spectrophotometer and an Agilent 2100 Bioanalyzer. Pyro-sequencing assays were designed and performed by BIOMARKER Company with both programs and assay result data supplied. mRNA was isolated by Oligo-dT magnetic beads from RNA, then the cDNA was synthesized using a QiaQuick PCR Extraction Kit (QIAGEN, USA). The cDNA library was constructed and sequenced by Illumina Hiseq 2500.

Differential Expressed Genes Analysis

Raw sequencing reads containing adaptors and low-quality (Q30 < 85%) were filtered. Then the remained reads were aligned to the genome of cucumber (9930 Version2) with TopHat2 (Kim et al., 2013), which mismatch was set as 2 and other parameters as the default value. FPKM (Fragments per Kilobase of transcript per Million fragments mapped) was used to detect the transcript abundance of each gene and estimate the expression values in all samples (Trapnell et al., 2010). Differentially expressed genes (DEGs) were identified according to the following criteria of |log2(fold change)| > 1 and false discovery rate (FDR) < 0.01 by DESeq2 (Love et al., 2014).

Genomic Methylation Analysis of Pericarp in Grafted Cucumber

In this study, the genome-wide methylation sequencing method was used according to the protocol described by Wang et al. (2015a), which was proved as a simple and scalable method. Total DNA of pericarp from self-rooted, grafted, and failed grafted cucumber with three biological replicates was extracted using a DNeasy Kit. According to the AFSM methylation technology, two restriction enzyme pairs of EcoRI-MspI and EcoRI-HpaII were used in the treatment of cucumber pericarp DNA samples in this study (Xia et al., 2014). The isoschizomers EcoRI was used as a frequent cutter while MspI and HpaII were used as rare cutters, and the methylation-susceptible sequences 59-CCGG and their methylation statuses were assessed by the AFSM method (Xia et al., 2014). The results of AFSM methylation were based on comparisons of the EcoRI-MspI and EcoRI-HpaII assembled sequences at the 59-CCGG sites using custom Perl scripts (http://afsmseq.sourceforge.net/) for each DNA sample analyzed by the China Golden Marker (Beijing) biotechnology company.

Expression Validation of Differentially Expressed and Methylated Genes

Quantitative real-time RT-PCR (qRT-PCR) technology was employed to verify the expression level of some candidate genes from RNA-seq results. The first-strand cDNA was amplified using a SYBR Green PCR Master Mix Kit (TaKaRa, Japan). Primer sequences for RT-PCR were designed on the "Quant Prime" website (http://quantprime.mpimp-golm.mpg.de/) (**Table S1**). qRT-PCR was conducted on an Applied Biosystems 7500 RT-PCR system (Thermo Fisher, USA) according to the manufacturer's instruction. The relative transcript levels of candidate genes and *TUA* (internal control) were calculated using the $2^{-\Delta\Delta Ct}$ method. Three biological and technical replicates were performed; the

differential gene expression level was determined by t-test ($p <$ 0.01).

Transcriptional Regulation Validation of *CsWIN1* in Pericarp Wax Biosynthesis

To investigate the transcriptional regulation of *CsWIN1*, cDNA of *CsWIN1* was amplified using specific primers (**Table S1**) and constructed into the vector pGADT7-Rec. The promoter sequences of four targeted genes (*CsCER1*, *CsCER1-1*, *CsCER4*, and *CsKCS1*) were amplified from genomic DNA and transformed into the pAbAi vector. Then the linearized library vectors pGADT7-Rec and pAbAi-HYT were co-transformed into the yeast strain Y1HGold. The transformed yeast cell with an empty pGADT7 vector was set as a negative control. The cultivation and transformation of yeast were carried out as described in the manufacturer's protocol. All primers used in this study are list in **Table S1**.

RESULTS

Observation of Pericarp Wax in Grafted Cucumber

The grafted cucumber exhibited a glossy pericarp with a light green color, in contrast to the self-rooted and failed grafted cucumber (**Figure 1A**). SEM (scanning electron microscope) observation revealed that the content and form of the wax was significantly different in grafted cucumber versus self-rooted cucumber. Fewer wax crystals were detected on the fruit surface, and the form of the wax was smaller and unbroken in grafted cucumber (**Figure 1B**). Wax crystals were more abundant in self-rooted and failed grafted cucumber and the wax exhibited a broken form, expressed in rods and tubes (**Figure 1B**). Cryosectioning that was conducted in both self-rooted and grafted cucumber showed that a thicker cuticle (15mm × 15mm) was clearly observed in self-rooted cucumber (**Figure 1C**).

Components Analysis of Pericarp Wax in Grafted Cucumber

The significantly reduced formation of wax powder in grafted cucumber prompted us to investigate the changes in chemical composition of cuticular wax using GC-mass spectrometry at the commercial maturity stage. We detected a significant decrease in the content of alkanes with length of 29 and 31 carbons (which are important components of wax) in grafted cucumber pericarp compared with the self-rooted and failed grafted cucumber (**Figure 2A**). The levels of these two alkanes components were decreased in the grafted versus self-rooted cucumber by approximately 50 and 30%, respectively. Grafted cucumber contained increased amounts of wax esters as compared to self-rooted cucumber ($p < 0.01$), especially these wax esters derived from C20 fatty acids, such wax ester C20:C22, wax ester C20:C24, wax ester C20:C26, wax ester C20:C28, and wax ester C20:29 (**Figure 2B**). Total wax ester levels were increased by approximately 15% in grafted cucumber compared to self-rooted cucumber. No noticeable changes in other alkanes and wax esters
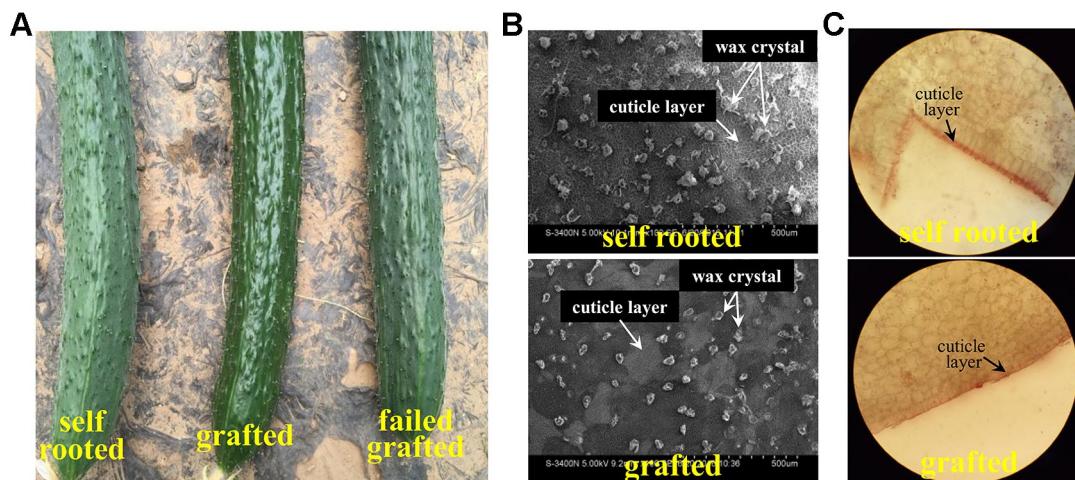
FIGURE 1 | Observation of pericarp appearance in grafted cucumber and in self-rooted cucumber. **Figure 1A** shows the pericarp of self-rooted, grafted, and failed grafted cucumber. The self-rooted and failed grafted pericarp samples were waxy while the grafted cucumber pericarp was glossy. **Figure 1B** displays the SEM observation of self-rooted and grafted cucumber pericarp. The small trichomes of self-rooted pericarp were broken, while those in the grafted pericarp were integrated. **Figure 1C** presents the frozen pathological examination. The cutin in self-rooted pericarp was found to be much thicker than in grafted cucumber.
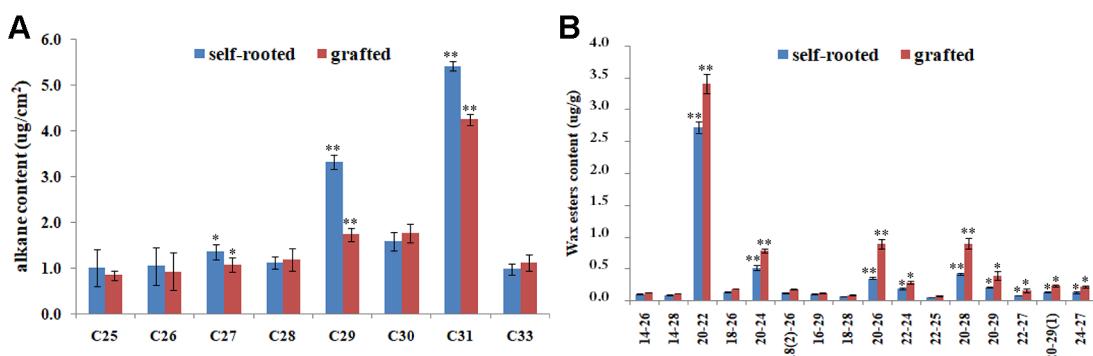
FIGURE 2 | Comparison of the composition of pericarp wax between self-rooted and grafted cucumber, analyzed by GC-MS. The major alkanes in **Figure 2A** were designated by carbon chain length in *x* axis and the amount of alkanes were expressed as µg/cm² pericarp area in *y* axis. The numbers in **Figure 2B** showed the carbon chain length of fatty acid and fatty alcohol in *x* axis, the corresponding wax esters amount were expressed as µg/g in *y* axis. Numbers in brackets indicated the amount of double bond. Error bars indicate standard errors. * and ** indicated significant differences of *t* test at $p < 0.05$ and $p < 0.01$, respectively.

were detected in the pericarp of grafted cucumber versus self-rooted cucumber.

Transcriptome Analysis of Pericarp Wax Biosynthesis After Grafting

Nine cucumber cDNA libraries from self-rooted, grafted, and failed grafted cucumber were sequenced on the Illumina HiSeq2500 platform; and 33.87 Gb clean reads were generated in total. Each library generated about 4.84 Gb clean reads; all Q30 reached more than 94.69% (**Table S2**). All of the clean sequencing data used in the present study are deposited in the NCBI Sequence Read Archive database (SRA accession: SRR10113591, SRR10113592, SRR10113593, SRR10113594, SRR 10113569, and SRR10113570). The sequencing reads from nine libraries showed a significant positive correlation (0.98) among the three

replications, indicating the reliable phenotype variations and high-quality sequencing data (**Figure S1**). All clean reads were aligned to Chinese long cucumber reference genome (9930 V2) and the mapping rate varied from 87.91 to 89.51% (**Table S2**). Then, all mapped reads were aligned with the databases of Swiss-Prot, KEGG, and GO using BLAST. Four hundred and twenty-four new genes were identified, 347 of which were annotated depended on the test for functional annotation.

Differentially expressed genes (DEGs) were identified by calculating the FPKM values among self-rooted, grafted, and failed grafted cucumber with the following criteria ($p < 0.05$, FDR < 0.01). Finally, a total of 384 DEGs were identified in this study (**Figure 3A**). Compared with self-rooted cucumber, 111 genes were upregulated while 82 genes were downregulated in grafted cucumber. In contrast to failed grafted cucumber, 118 genes were upregulated and 111 genes were downregulated

FIGURE 3 | Transcriptome dataset and expression validation of differentially expressed genes affected by grafting at commercial mature stage. **Figure 3A** presents a Venn diagram of differentially expressed genes (DEGs) obtained from the self-rooted, grafted, and failed grafted cucumber pericarp. **Figure 3B** shows the qRT-PCR validation (three technical replicates per biological replicate) of the 10 differentially expressed genes involved pericarp wax biosynthesis. Error bars indicate the standard deviations. * and ** indicated significant differences of t test at $p < 0.05$ and $p < 0.01$, respectively.

in grafted cucumber. Above all, 68 genes were found to have significant differential expression levels in grafted cucumber compared with both self-rooted and failed grafted cucumber (**Table S3**), indicating that these genes could play a role in the pericarp wax biosynthesis of grafted cucumber. Based on the gene functions involved in wax biosynthesis and highest fold changes in grafted cucumber, 10 of the 68 DEGs were considered to be important candidate genes affecting pericarp wax formation (**Table 1**). These 10 DEGs had different functions associated with wax metabolism, including four DEGs involved in wax biosynthesis, three involved in wax transportation, and three AP2/ERF transcriptional factors that may be involved in the regulation of the gene expression of wax biosynthesis.

Genomic Methylation Analysis of Pericarp Wax Synthesis After Grafting

A total of 4,247 methylated genes were detected in this study, 1,184 of which were specifically observed in grafted cucumber (as compared to self-rooted and failed grafted cucumber).

Among these genes, the DNA methylated regions in the UTR, exon, intron, and promoter regions were distributed as 4, 48.8, 32.7, and 14.2% in the grafting-induced methylated genes, respectively. Twenty grafting-induced methylated genes were observed in the profile of differentially expressed genes in the grafted, but not in the un-grafted cucumber pericarp. These were identified as critical candidate genes affecting the changes in appearance of the scion pericarp in cucumber (**Table 2**). In contrast, we observed 2,521 methylated genes in the negative control but not in grafted treatment, these genes were assumed as the demethylated genes after graft, which included 43 DEGs based on transcriptome analysis (**Table S4**).

The methylation sequencing results indicated that the intron region of *Csa3G878210* possesses a CCGG type methylation in grafted cucumber which was not detected in both negative control treatments. Furthermore, the expression of *Csa3G878210* was up-regulated in grafted cucumber compared with the self-rooted and failed grafted cucumber treatment. This gene is an AP2/ERF transcriptional factor and the homologous gene of the wax biosynthesis regulator gene *WIN1* (AP2/ERF gene)

TABLE 1 | The 10 differentially expressed genes involved in wax biosynthesis.

Gene category	Gene ID	FPKM			Gene Annotation
		SR[a]	G	FG	
Wax	Csa3G127750	3.5	57.6	3.0	CER1; contains fatty acid hydroxylase
	Csa6G079750	91.8	205.0	83.9	CER1 protein, putative, expressed
	Csa6G151810	6.1	28.7	9.0	CER4, fatty acyl-CoA reductase
	Csa6G302180	9.3	23.0	11.3	KCS1, 3-ketoacyl-CoA synthase
ABC/LTP	Csa3G446120	2.4	9.5	4.1	ABC transporter G family member
	Csa3G027200	14.5	42.1	33.8	14 kDa proline-rich protein DC2.15, putative
	Csa4G017140	154.5	214.3	104.6	PVR3-like protein
AP2/ERF	Csa6G496390	2.1	0.9	2.3	AP2-like ethylene-responsive transcription factor
	Csa3G878210	8.9	24.7	5.8	Ethylene-responsive transcription factor 1a
	Csa5G167110	6.7	20.8	15.5	Ethylene-responsive transcription factor

[a]SR, self-rooted; G, grafted; FG, failed grafted.

TABLE 2 | List of 20 DEGs with DNA methylation in grafted cucumber.

Gene ID	FPKM			Gene Annotation
	SR[a]	G	FG	
Csa3G878210	24.7	6.7	15.5	Ethylene-responsive transcription factor 1a
Csa3G446120	9.5	2.4	4.1	ABC transporter G family member
Csa6G488880	15.6	10.1	7.4	Putative receptor-like protein kinase
Csa6G446400	10.8	7.2	4.8	Transcription factor, putative
Csa2G352420	38.4	7.9	26.3	Zinc finger CCCH domain-containing protein
Csa3G739050	23.8	11.4	17.2	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MOK16
Csa1G023030	35.8	29.8	16.4	Unknown protein
Csa3G817740	0.9	0.8	0.1	Probable exocyst complex component 6
Csa6G237600	109.1	52.3	104.4	Basic 7S globulin
Csa5G517100	5.5	5.3	2.6	Actin 4
Csa3G150000	117.1	50.2	113.1	Xyloglucan-specific endoglucanase inhibitor protein
Csa6G504490	6	5.9	2.1	ATP-dependent zinc metalloprotease FtsH 2
Csa7G048060	6.1	1.3	6.5	Plant-specific domain TIGR01615 family protein
Csa1G046270	3.5	5.1	7.3	Receptor protein kinase, putative
Csa5G199270	7.7	14.5	16.7	Aquaporin
Csa1G084320	116.9	297.7	226.9	Histone H4
Csa2G215520	2.1	5.8	1.6	Sulfate adenylyl transferase
Csa2G409480	2.3	5.6	6.6	Zinc finger family protein
Csa4G002500	0.4	1.2	1.1	Cellulose synthase-like protein
Csa2G196890	2	8.9	3.2	Sieve element occlusion protein 1

[a]*Abbreviation was same as in **Table 1**.*

in *Arabidopsis*. Hence, this grafting-induced methylated gene (*Csa3G878210*) was named *CsWIN1*, which may play a critical role in regulation of wax biosynthesis in cucumber, especially in brightening the appearance of grafted pericarp, resulting in glossy cucumbers.

Expression Validation of Wax Biosynthesis-Related Genes Affected by Grafting

In this study, the 10 obtained wax biosynthesis genes in transcriptome profiles showed considerable differences in gene expression levels between self-rooted and grafted cucumber pericarp. To confirm this results, qRT-PCR evaluation was performed in the self-rooted and grafted cucumber pericarp. There were eight genes (*CsCER1*, *CsCER1-1*, *CsCER4*, *CsKCS1*, *CsABC*, *Csa4G017140*, *Csa6G496390*, and *CsWIN1*) that were significantly ($p < 0.01$) upregulated in grafted cucumber, while one ERF gene (*Cs6G496390*) was downregulated. Thus, these results validated the transcriptome data (**Figure 3B**). However, one gene (*CsLTP*) showed a difference in expression level without significance. This validation result was consistent with the transcriptome profiles. Most of the wax biosynthesis-related genes were upregulated in the grafted cucumber pericarp (**Table 1**).

## Transcription Regulation Validation of *CsWIN1* in Pericarp Wax Biosynthesis

In this study, five well-investigated AP2/ERF TFs in the regulation of wax biosynthesis in *Arabidopsis* were analyzed in the cucumber genome, and three homologue genes were obtained among the 68 DEGs identified in the transcriptome data, including the key *CsWIN*1 gene, which was specifically methylated by grafting in the cucumber pericarp, as well as two

ERF genes. A phylogenetic tree aligned using the neighbor-joining (NJ) method through MEGA7 software were constructed and *CsWIN1* was confirmed to be closely related to the key wax biosynthesis regulator *WIN1* in *Arabidopsis* (**Figure 4A**) (Shi et al., 2011). In this study, the transcription regulation validation of *CsWIN1* was examined in the regulation of wax biosynthesis genes by a yeast one-hybrid system (Y1H). The promoter sequences of two targeted genes (*CsCER1* and *CsCER4*) including the element "GCCGGC" were amplified from genomic DNA and transformed into the pAbAi vector. The yeast cells that grew on the Y1H Gold-carrying PGADT7-*CsWIN1* vectors and the pBait-AbAi vectors could be grown in SD/-Leu/AbA 100 medium, and the positive clones increased gradually with the increase of the dilution concentration. This result indicated that pGADT7-*CsWIN1* harbors a transcriptional activation domain preferentially binding to the "GCCGGC" box of the *CsCER1* and *CsCER4* promoter (**Figure 4B**). The key grafting-induced regulator *CsWIN1* transcriptionally activates the expression of target wax biosynthesis genes, thus regulating the content and composition of pericarp wax in cucumber, resulting in the brightening and increased glossiness of the pericarp.

DISCUSSION

The Unique Character of Cucumber Pericarp, as Well as Appearance After Grafting

In *Arabidopsis*, the wax appearance was crystal-like in the stem and leaves as well as in siliques (Haslam and Kunst, 2012). The chemical composition and content of cuticular wax is different among species and tissues (Kunst and Samuels, 2009). The cuticle wax responds to external environment stresses
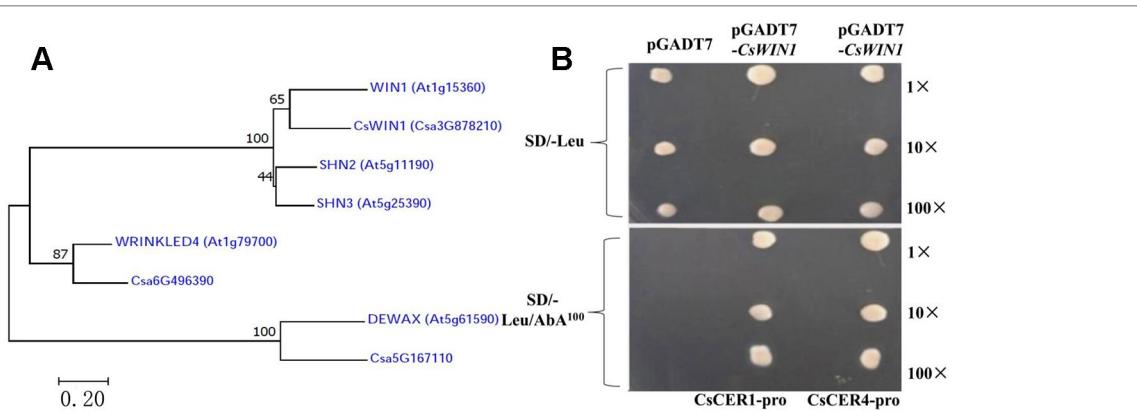
FIGURE 4 | Phylogenetic analysis and Y1H transcriptional test of *CsWIN1* in the regulation of wax biosynthesis. **Figure 4A** displays the key wax regulator genes in *Arabidopsis* and their homologues in cucumber. **Figure 4B** shows the transcriptional binding test of *CsWIN1* with *CsCER1* and *CsCER4* promoters by Y1H.

and is a physical and chemical barrier on the outer surfaces of terrestrial plants (Yeats and Rose, 2013; Ziv et al., 2018). The wax of cucumber fruit was observed as small balls or trichomes, quite different from that in *Arabidopsis* and other crops like wheat (Hen-Avivi et al., 2016). The small balls or trichomes in the glossy cucumber are integrated, while as those in waxy cucumbers appear broken and exist as fragmentary forms that are uniformly distributed on the surface of the cucumber fruit (**Figure 1B**). Previous studies also found that the chemical components of wax was different between *Arabidopsis* and cucumber, which phenols and alkenes were uniquely detected, while the ketones was unable to identified in cucumber (Wang et al., 2015b; Wang et al., 2015c).

In this study, we observed that after grafting with the rootstock of "Jingxinzhen 6", the cucumber pericarp on the scion became glossy, as compared with the waxy appearance of the self-rooted cucumber (**Figure 1A**). The metabolite analysis indicates that there is a lower alkane content, but more wax esters were accumulated in the grafted cucumber.

Several studies have shown that the surface of cucumber fruit is more affected by silica, which is a natural powder covering on the surface of cucumber fruits, rather than the cuticular wax (Mitani et al., 2011). This finding was also investigated in the grafting system by comparing the effects of different rootstocks—*C. moschata* and *Cucurbita ficifilia*. The former pumpkin rootstock enhanced the glossy appearance of cucumber fruit, whereas the latter rootstock did not (Seki and Hotta, 1997). Another study showed that one amino acid mutation in the silicon influx transporter gene (AQP family genes) could lead to less silicon uptake in the soil, resulting in a glossy cucumber fruit when grafted onto *C. moschata* (Mitani et al., 2011). Here, we observed that four AQP family genes were all downregulated by grafting. These results are consistent with these previous studies (**Table S3**). Combined with the fact that the cuticular wax composition was significantly changed after grafting in this study, we hypothesized that there were two regulation pathways (wax biosynthesis and Si absorption) through which the rootstock

affects the scion in grafting systems, though further research is required to validate this.

Regulated Wax Biosynthesis Genes in the Pericarp After Grafting in Cucumber

Grafting is an important way to improve plant growth, stress tolerance, and fruit quality, and has been widely used in commercial horticultural crop production (Zahaf et al., 2012; Liu et al., 2016). The molecular mechanisms regulating plant growth *via* grafting have been investigated by several groups (Kyriacou et al., 2017), for example using high-throughput sequencing in watermelon (Liu et al., 2013), tomato (Estan et al., 2005), apple (An et al., 2018), and grapevine (Pagliarani et al., 2017). These studies showed that grafting affects differential expression of genes in grafted plants. miRNAs are also exchanged between the scion and rootstock in grafted watermelon, which may regulate the growth and development of the scion (Cookson and Ollat, 2013; Liu et al., 2013). A genomic DNA methylation analysis was performed in cucumber and melon scions by a grafting system using cucurbitaceous inter-grafting (Avramidou et al., 2015). Recently, some research identified many DEGs involved in various metabolic processes in grafted cucumber and verified by qRT-PCR (Miao et al., 2019a; Miao et al., 2019b). In this study, we observed that pericarp phenotype of cucumber grafted on pumpkin rootstock is distinct from these of the failed grafted cucumber and self-rooted cucumber. A previous study (Jun et al., 2016) also reported that the pericarp wax powder content in the self-grafted cucumber had no difference with self-rooted cucumber.

The present study identified 68 significantly DEGs in grafted cucumber and 10 of them were annotated in the wax biosynthesis pathway (**Table 1**). These 10 DEGs associated with wax biosynthesis were validated by qRT-PCR (**Figure 3B**). These DEGs may also function in VLCFAs biosynthesis and wax ester biosynthesis, as well as in the transportation of cuticular wax inside or outside of the cell in cucumber scion (McFarlane et al., 2010; Kim et al., 2012).

This study also examined the genomic methylation of cucumber scion grafted in pumpkin rootstock based on a genome-wide methylation sequencing method (Wang et al., 2015a), which was proved as a simple and scalable method. Although it may not detect all methylated genes in the genome because of limited sites by enzymes and the low coverage sequencing, it was a cost-effectively method in identifying candidate genes and had been used in many researches (Mastan et al., 2012; Rathore et al., 2014; Tang et al., 2014; Xia et al., 2014). Here, 20 key genes were methylated and the association analysis between the transcriptome and methylation identified one key wax biosynthesis regulator—*CsWIN1*, homologous to the ERF family members *WIN*, *WR4*, and *DEWAX* in *Arabidopsis*. The close phylogenetic relationship between *CsWIN1* and *WIN1* provides confirmation evidence that the ERF gene *CsWIN1* is an important regulator in wax biosynthesis. Moreover, this gene also affects expression of other genes (*CsCER1* and *CsCER4)* in the wax biosynthesis pathway, in addition to the downstream transporter gene *CsABC* in the cucumber scion.

Potential Regulation Model of *CsWIN1* Affecting Pericarp Wax Biosynthesis

The key *CsWIN1* gene was observed in the transcriptome data and in the genomic DNA methylation profile. Furthermore, it is homologous to a key wax biosynthesis regulator in *DEWAX* in the model plant *Arabidopsis* (**Table S2**; **Figure 4A**). The expression of *CsWIN1* is positively correlated with four cuticular wax biosynthesis genes (*CsCER1*, *CsCER1-1*, *CsCER4*, and *CsKCS1*) that were upregulated after grafting on the pumpkin rootstock (**Figure 3**). In addition, the potential wax

transporter genes (ABC/LTP transporter) were upregulated and co-expressed with the key regulator *CsWIN1* in the graft treatment (**Table S2**). Based on these results, we hypothesized that the key regulator *CsWIN1* affected by grafting could be a master switch that transcriptionally regulates the expression of the wax biosynthesis pathway genes in response to grafting, specifically activating *CsCER1*, *CsCER1-1*, *CsCER4*, and *CsKCS1* as well as transportation members of the ABC/LTP transporter family in the scion. The Y1H experiment validated that the key regulator could bind to the promoter region of *CsCER1* and Cs*CER4*, illustrating the feasibility of transcriptional regulation (**Figure 4B**). Moreover, the metabolite dataset showed that the content composition of cuticular wax the cucumber scion was significantly affected by grafting on pumpkin.

Following grafting, fewer alkanes and more wax esters were accumulated on the pericarp and the wax balls were not broken. Therefore, we propose a possible model of *CsWIN1* in the regulation of wax biosynthesis in grafted cucumber (**Figure 5**). *CsWIN1* was methylated and upregulated in grafted cucumber, and then transcriptionally activated the expression of wax biosynthesis genes *CsCER1* and *CsCER4*, and may regulate the expression of transporter gene *CsABC*, resulting in the biosynthesis and transportation of more wax esters into the pericarp. This process makes the small trichomes less prone to breaking in the grafted cucumber, as opposed to being easily broken in the self-rooted cucumber. This model shows a molecular regulation pathway of wax biosynthesis in cucumber and could provide a reference to other crops. However, the *in vitro* function of *CsWIN1* still needs to be validated by transformation experiments. The pumpkin's genetic factors (genes, RNAs and proteins) that affect the cucumber pericarp are unknown and could be identified in future studies.
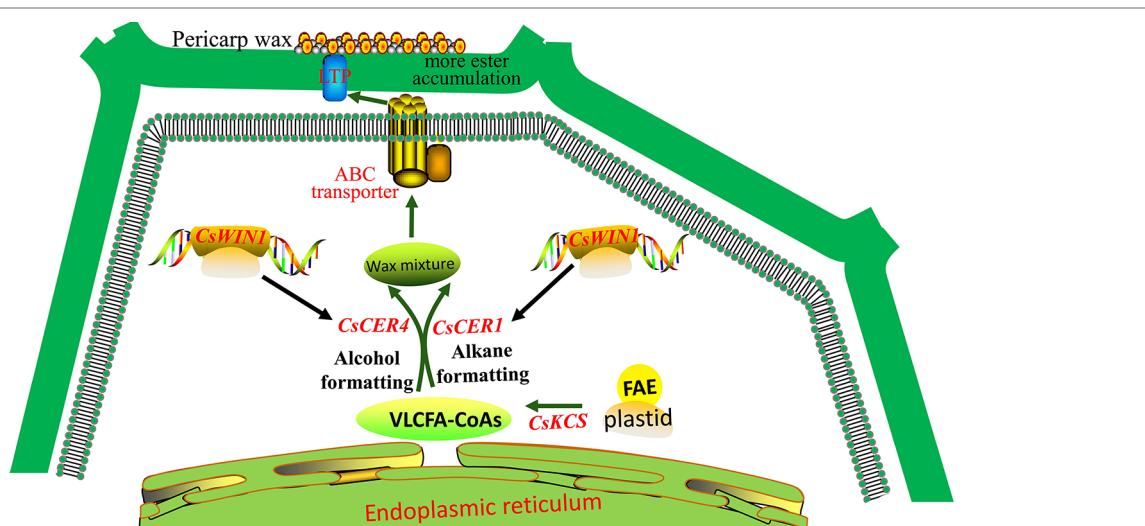
FIGURE 5 | The potential regulation model of *CsWIN1* in wax biosynthesis. *CsWIN1* was methylated and upregulated by a grafting test. It then transcriptionally activated the expression of wax biosynthesis genes *CsCER1* and *CsCER4*, and may regulate the expression of transporter gene *CsABC*, resulted in the biosynthesis and transportation of more wax esters into the pericarp. This make the small trichomes not so fragile, so they do not break up in the grafted cucumber, as opposed to being easily broken in the self-rooted cucumber.

DATA AVAILABILITY STATEMENT

All datasets for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

CW and YX designed this research. JZ and YY performed the grafting experiment, JY and JZ analyzed the bioinformatics data, and JL, JZ, and XZ performed the validation test. JZ, CW, and YX wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01564/full#supplementary-material

FIGURE S1 | Heat map of all DEGs in self-rooted (SR), grafted (G), and failed grafted (FG) cucumber in three biological replications.

TABLE S1 | All primer sequences used in this study.

TABLE S2 | The basic analysis of RNA-seq reads.

TABLE S3 | FPKM values of differentially expressed genes affected by grafting.

TABLE S4 | List of 43 DEGs with DNA demethylation in grafted cucumber.

REFERENCES

An, N., Fan, S., Yang, Y., Chen, X., Dong, F., Wang, Y., et al. (2018). Identification and Characterization of miRNAs in Self-Rooted and Grafted Malus Reveals Critical Networks Associated with Flowering. *Int. J. Mol. Sci.* 19, 2384. doi: 10.3390/ijms19082384

Avramidou, E., Kapazoglou, A., Aravanopoulos, F. A., Xanthopoulou, A., Ganopoulos, I., Tsaballa, A., et al. (2015). Global DNA methylation changes in Cucurbitaceae inter-species grafting. *Crop Breed. Appl. Biotechnol.* 15, 112–116. doi: 10.1590/1984-70332015v15n2n20

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J. D., et al. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that Arabidopsis ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* 24, 3106–3118. doi: 10.1105/tpc.112.099796

Blanc, C., Coluccia, F., L'Haridon, F., Torres, M., Ortiz-Berrocal, M., Stahl, E., et al. (2018). The Cuticle Mutant eca2 Modifies Plant Defense Responses to Biotrophic and Necrotrophic Pathogens and Herbivory Insects. *Mol. Plant-Microbe Interact. : MPMI* 31, 344–355. doi: 10.1094/MPMI-07-17-0181-R

Broun, P., Poindexter, P., Osborne, E., Jiang, C. Z., Riechmann, J. L., and Chrispeels, M. J. (2004). WIN1, a Transcriptional Activator of Epidermal Wax Accumulation in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4706–4711. doi: 10.1073/pnas.0305574101

Buschhaus, C., Hager, D., and Jetter, R. (2015). Wax Layers on Cosmos bipinnatus Petals Contribute Unequally to Total Petal Water Resistance. *Plant Physiol.* 167, 80. doi: 10.1104/pp.114.249235

Cookson, S. J., and Ollat, N. (2013). Grafting with rootstocks induces extensive transcriptional re-programming in the shoot apical meristem of grapevine. *BMC Plant Biol.* 13, 147. doi: 10.1186/1471-2229-13-147

Estan, M. T., Martinez-Rodriguez, M. M., Perez-Alfocea, F., Flowers, T. J., and Bolarin, M. C. (2005). Grafting raises the salt tolerance of tomato through limiting the transport of sodium and chloride to the shoot. *J. Exp. Bot.* 56, 703–712. doi: 10.1093/jxb/eri027

Feng, J. X., Liu, D., Pan, Y., Gong, W., Ma, L. G., Luo, J. C., et al. (2005). An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EREBP transcription factor gene family. *Plant Mol. Biol.* 59, 853–868. doi: 10.1007/s11103-005-1511-0

Gan, L., Wang, X., Cheng, Z., Liu, L., Wang, J., Zhang, Z., et al. (2016). Wax crystal-sparse leaf 3 encoding a β-ketoacyl-CoA reductase is involved in cuticular wax biosynthesis in rice. *Plant Cell Rep.* 35, 1687–1698. doi: 10.1007/s00299-016-1983-1

Go, Y. S., Kim, H., Kim, H. J., and Mi, C. S. (2014). Arabidopsis Cuticular Wax Biosynthesis Is Negatively Regulated by the DEWAX Gene Encoding an AP2/ERF-Type Transcription Factor. *Plant Cell* 26, 1666–1680. doi: 10.1105/tpc.114.123307

Gutterson, N., and Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr. Opin. Plant Biol.* 7, 465–471. doi: 10.1016/j.pbi.2004.04.007

Haslam, T. M., and Kunst, L. (2012). Arabidopsis ECERIFERUM2 is a component of the fatty acid elongation machinery required for fatty acid extension to exceptional lengths. *Plant Physiol.* 160, 1164. doi: 10.1104/pp.112.201640

Hen-Avivi, S., Savin, O., Racovita, R. C., Lee, W.-S., Adamski, N. M., Malitsky, S., et al. (2016). A metabolic gene cluster in the wheat W1 and the barley Cer-cqu loci determines β-diketone biosynthesis and glaucousness. *Plant Cell* 28, 1440–1460. doi: 10.1105/tpc.16.00197

Huang, S., Ruiqiang, L., Zhonghua, Z., Li, L., Xingfang, G., Wei, F., et al. (2009). The genome of the cucumber, Cucumis sativus L. *Nat. Genet.* 41, 1275. doi: 10.1038/ng.475

Huang, P. Y., Catinot, J., and Zimmerli, L. (2016). Ethylene response factors in Arabidopsis immunity. *J. Exp. Bot.* 67, 1231–1241. doi: 10.1093/jxb/erv518

Jun, H. U., Xiang, C., Wang, C., Duan, Y., and Chen, H. (2016). *Effect of Rootstock on Silicon and Fruit Surface Wax Powder Content in Grafted Cucumber* (Northern Horticulture).

Kim, H., Lee, S. B., Kim, H. J., Min, M. K., Hwang, I., and Suh, M. C. (2012). Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in Arabidopsis thaliana. *Plant Cell Physiol.* 53, 1391–1403. doi: 10.1093/pcp/pcs083

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36

Kunst, L., and Samuels, L. (2009). Plant cuticles shine: advances in wax biosynthesis and export. *Curr. Opin. Plant Biol.* 12, 721–727. doi: 10.1016/j.pbi.2009.09.009

Kyriacou, M., Rouphael, Y., Colla, G., Zrenner, R., and Schwarz, D. (2017). Vegetable Grafting: The Implications of a Growing Agronomic Imperative for Vegetable Fruit Quality and Nutritive Value. *Front. In Plant Sci.* 8, 741. doi: 10.3389/fpls.2017.00741

Lam, S. M., Tong, L., Reux, B., Lear, M. J., Wenk, M. R., and Shui, G. (2013). Rapid and sensitive profiling of tear wax ester species using high performance liquid chromatography coupled with tandem mass spectrometry. *J. Chromatogr. A* 1308, 166–171. doi: 10.1016/j.chroma.2013.08.016

Lee, S. B., and Mi, C. S. (2015). Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in Arabidopsis. *Plant Cell Physiol.* 56, 48. doi: 10.1093/pcp/pcu142

Lee, S. B., Jung, S. J., Go, Y. S., Kim, H. U., Kim, J. K., Cho, H. J., et al. (2009). Two Arabidopsis 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but

differentially controlled by osmotic stress. *Plant J. : Cell Mol. Biol.* 60, 462–475. doi: 10.1111/j.1365-313X.2009.03973.x

Li, C. W., Su, R. C., Cheng, C. P., Sanjaya, You, S. J., Hsieh, T. H., et al. (2011). Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol.* 156, 213–227. doi: 10.1104/pp.111.174268

Li, L., Du, Y., He, C., Dietrich, C. R., Li, J., Ma, X., et al. (2019). Maize glossy6 is involved in cuticular wax deposition and drought tolerance. *J. Exp. Bot.* 70 (12), 3089–3099. doi: 10.1093/jxb/erz131

Liu, N., Yang, J., Guo, S., Xu, Y., and Zhang, M. (2013). Genome-wide identification and comparative analysis of conserved and novel microRNAs in grafted watermelon by high-throughput sequencing. *PloS One* 8, e57359. doi: 10.1371/journal.pone.0057359

Liu, J., Li, J., Su, X., and Xia, Z. (2014). Grafting improves drought tolerance by regulating antioxidant enzyme activities and stress-responsive gene expression in tobacco. *Environ. Exp. Bot.* 107, 173–179. doi: 10.1016/j.envexpbot.2014.06.012

Liu, N., Yang, J., Fu, X., Zhang, L., Tang, K., Guy, K. M., et al. (2016). Genome-wide identification and comparative analysis of grafting-responsive mRNA in watermelon grafted onto bottle gourd and squash rootstocks by high-throughput sequencing. *Mol. Genet. Genomics* 291, 621–633. doi: 10.1007/s00438-015-1132-5

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

Mastan, S. G., Rathore, M. S., Bhatt, V. D., Yadav, P., and Chikara, J. (2012). Assessment of changes in DNA methylation by methylation-sensitive amplification polymorphism in Jatropha curcas L. *Subjected Salinity Stress Gene* 508, 125–129. doi: 10.1016/j.gene.2012.07.063

McFarlane, H. E., Shin, J. J., Bird, D. A., and Samuels, A. L. (2010). Arabidopsis ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. *Plant Cell* 22, 3066–3075. doi: 10.1105/tpc.110.077974

Miao, L., Di, Q., Sun, T., Li, Y., Duan, Y., Wang, J., et al. (2019a). Integrated Metabolome and Transcriptome Analysis Provide Insights into the Effects of Grafting on Fruit Flavor of Cucumber with Different Rootstocks. *Int. J. Mol. Sci.* 20 (14), 3592. doi: 10.3390/ijms20143592

Miao, L., Qin, X., Gao, L., Li, Q., Li, S., He, C., et al. (2019b). Selection of reference genes for quantitative real-time PCR analysis in cucumber (Cucumis sativus L.), pumpkin (Cucurbita moschata Duch.) and cucumber-pumpkin grafted plants. *PeerJ* 7, e6536. doi: 10.7717/peerj.6536

Mitani, N., Yamaji, N., Ago, Y., Iwasaki, K., and Ma, J. F. (2011). Isolation and functional characterization of an influx silicon transporter in two pumpkin cultivars contrasting in silicon accumulation. *Plant J. Cell Mol. Biol.* 66, 231–240. doi: 10.1111/j.1365-313X.2011.04483.x

Pagliarani, C., Vitali, M., Ferrero, M., Vitulo, N., Incarbone, M., Lovisolo, C., et al. (2017). Accumulation of MicroRNAs Differentially Modulated by Drought Is Affected by Grafting in Grapevine. *Plant Physiol.* 173, pp.01119–02016. doi: 10.1104/pp.16.01119

Park, C. S., Go, Y. S., and Suh, M. C. (2016). Cuticular wax biosynthesis is positively regulated by WRINKLED4, an AP2/ERF-type transcription factor, in Arabidopsis stems. *Plant J.* 88, 257–270. doi: 10.1111/tpj.13248

Rathore, M. S., Mastan, S. G., and Agarwal, P. K. (2014). Evaluation of DNA methylation using methylation-sensitive amplification polymorphism in plant tissues grown in vivo and in vitro. *Plant Growth Regul.* 75, 11–19. doi: 10.1007/s10725-014-9926-8

Samuels, A. L., Glass, A. D. M., Ehret, D. L., and Menzies, J. G. (1993). The Effects of Silicon Supplementation on Cucumber Fruit: Changes in Surface Characteristics. *Ann. Bot.* 72, 433–440. doi: 10.1006/anbo.1993.1129

Seki, M., and Hotta, Y. (1997). *Effect of bloomless stock cultivar on the growth and mineral uptake of cucumber [Cucumis sativus] plants* ((Japan): Research Bulletin of the Aichi-ken Agricultural Research Cente).

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

Shen, Q. H., Ren, Z. H., Wei, M., Li, Y., Wang, X. F., and Cui, J. (2015). Effects of silicon supply on silicon absorption and bloom formation on fruit surface of cucumber grafted on different rootstocks. *Acta Hortic.* 1086, 231–236. doi: 10.17660/ActaHortic.2015.1086.29

Shi, J. X., Malitsky, S., De, O. S., Branigan, C., Franke, R. B., Schreiber, L., et al. (2011). SHINE transcription factors act redundantly to pattern the archetypal surface of Arabidopsis flower organs. *PloS Genet.* 7, e1001388. doi: 10.1371/journal.pgen.1001388

Suh, M. C., Samuels, A. L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., et al. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. *Plant Physiol.* 139, 1649–1665. doi: 10.1104/pp.105.070805

Tang, X.-M., Tao, X., Wang, Y., Ma, D.-W., Li, D., Yang, H., et al. (2014). Analysis of DNA methylation of perennial ryegrass under drought using the methylation-sensitive amplification polymorphism (MSAP) technique. *Mol. Genet. Genomics* 289, 1075–1084. doi: 10.1007/s00438-014-0869-6

Tian, G.-L., Zhang, S.-P., Song, Z.-C., Zhang, S., Cui, J.-Y., Miao, H., et al. (2015). Genetic Analysis and QTL Mapping of Wax Powder on the Surface of Cucumber Fruit. *Sci. Agric. Sin.* 2015, 10.

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

Wang, S., Lv, J., Zhang, L., Dou, J., Sun, Y., Li, X., et al. (2015a). MethylRAD: a simple and scalable method for genome-wide DNA methylation profiling using methylation-dependent restriction enzymes. *Open Biol.* 5, 150130. doi: 10.1098/rsob.150130

Wang, W., Liu, X., Gai, X., Ren, J., Liu, X., Cai, Y., et al. (2015b). Cucumis sativus L. WAX2 Plays a Pivotal Role in Wax Biosynthesis, Influencing Pollen Fertility and Plant Biotic and Abiotic Stress Responses. *Plant Cell Physiol.* 56, 1339–1354. doi: 10.1093/pcp/pcv052

Wang, W., Zhang, Y., Xu, C., Ren, J., Liu, X., Black, K., et al. (2015c). Cucumber ECERIFERUM1 (CsCER1), which influences the cuticle properties and drought tolerance of cucumber, plays a key role in VLC alkanes biosynthesis. *Plant Mol. Biol.* 87, 219–233. doi: 10.1007/s11103-014-0271-0

Xia, Z., Zou, M., Zhang, S., Feng, B., and Wang, W. (2014). AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. *Sci. Rep.* 4, 7300. doi: 10.1038/srep07300

Yeats, T. H., and Rose, J. K. C. (2013). The Formation and Function of Plant Cuticles. *Plant Physiol.* 163, 5–20. doi: 10.1104/pp.113.222737

Zahaf, O., Blanchet, S., De Zélicourt, A., Alunni, B., Plet, J., Laffont, C., et al. (2012). Comparative transcriptomic analysis of salt adaptation in roots of contrasting Medicago truncatula genotypes. *Mol. Plant* 5, 1068–1081. doi: 10.1093/mp/sss009

Ziv, C., Zhao, Z., Gao, Y. G., and Xia, Y. (2018). Multifunctional Roles of Plant Cuticle During Plant-Pathogen Interactions. *Front. Plant Sci.* 9, 1088. doi: 10.3389/fpls.2018.01088

Inheritance and Quantitative Trait Locus Mapping of *Fusarium* Wilt Resistance in Cucumber

Jingping Dong, Jun Xu, Xuewen Xu, Qiang Xu and Xuehao Chen*

School of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China

Fusarium wilt (FW) is a very serious soil-borne disease worldwide, which usually results in huge yield losses in cucumber production. However, the inheritance and molecular mechanism of the response to FW are still unknown in cucumber (*Cucumis sativus* L.). In this study, two inbred cucumber lines Superina (P_1) and Rijiecheng (P_2) were used as the sensitive and resistant lines, respectively. A mixed major gene plus polygene inheritance model was used to analyze the resistance to FW in different generations of cucumber, namely, P_1, P_2, F_1 ($P_1 \times P_2$), B_1, and B_2, obtained by backcrossing F_1 plants with Superina (B_1) or Rijiecheng (B_2), and F_2, obtained by self-crossing the F_1 plants. After screening 18 genetic models, we chose the E-1 model, which included two pairs of additive-dominance-epistatic major genes and additive-dominance polygenes, as the optimal model for resistance to FW on the basis of fitness tests. The major effect quantitative trait locus (QTL) *fw2.1* was detected in a 1.91-Mb-long region of chromosome 2 by bulked-segregant analysis. We used five insertion/deletion markers to fine-map the *fw2.1* to a 0.60 Mb interval from 1,248,093 to 1,817,308 bp on chromosome 2 that contained 80 candidate genes. We also used the transcriptome data of Rijiecheng inoculated with *Fusarium oxysporum* f. sp. *cucumerinum* (Foc) to screen the candidate genes. Twelve differentially expressed genes were detected in *fw2.1*, and five of them were significantly induced by FW. The expression levels of the five genes were higher in FW-resistant Rijiecheng inoculated with Foc than in the control inoculated with water. Our results will contribute to a better understanding of the genetic basis of FW resistance in cucumber, which may help in breeding FW-resistant cucumber lines in the future.

Keywords: cucumber, *Fusarium* wilt, inheritance, quantitative trait locus fine mapping, resistance genes

INTRODUCTION

Cucumber (*Cucumis sativus* L.) is an economically important vegetable crop that ranks fourth in vegetable production worldwide. In 2017, the cultivated area in China was 1.24 million hectares with an annual output of 64.88 million tons, respectively accounting for 54.4% and 77.5% of the world total 2017 (http://faostat3.fao.org/). In the cucumber production season, *Fusarium* wilt (FW), downy mildew, and powdery mildew are the three main diseases found in China (Cao et al., 2007). FW of cucumber occurs more readily and seriously under a continuous cropping system, with incidences ranging from 30% to 90%, leading to huge loss of cucumber output (Pu et al., 2011; Zhou and Wu, 2012).

FW caused by *Fusarium oxysporum* f. sp. *cucumerinum* Owen (Owen, 1955), which is a forma specialis that infects the vascular bundles of cucumber, leads to necrotic lesions on the stem base, foliar wilting and eventually whole plant wilt, and even death (Zhou et al., 2008; Ye et al., 2004). Until now, there are no methods to effectively control the occurrence and harm of cucumber FW (Ye et al., 2004; Yuan et al., 2003).

Understanding the inheritance of FW resistance is an important step in developing resistant breeding resources and in breeding resistant varieties. The inheritance of cucumber resistance to FW has been studied for a long time, but the results are inconsistent (Pierce and Wehner, 1990; Vakalounakis, 1993). Vakalounakis, (2015) found that cucumber resistance to FW was controlled by a single gene, whereas others have suggested that cucumber FW resistance was regulated by multiple genes (Liu et al., 2003; Wang, 2005). To date, the inheritance of cucumber FW resistance remains poorly understood. Disease-resistant traits are quantitative traits that are controlled by multiple genes, which are generally located in multiple effect quantitative trait loci (QTLs; Hiroki et al., 2013; Falconer and Mackay, 1996). Bulked-segregant analysis (BSA) is an effective method for identifying DNA markers tightly linked to FW resistance (Giovannoni et al., 1991; Michelmore et al., 1991). For example, Li et al. (2018) mapped powdery mildew resistance genes to a region of chromosome 1A of wheat, and Zhang et al. (2018) confirmed that resistance to downy mildew and powdery mildew shared common candidate intervals on chromosome 5 of cucumber (Zhang et al., 2018) by BSA. DNA samples from soybean with opposite phenotypes were subjected to BSA to detect DNA markers that exhibited differences between the two different samples to identify the QTL (Mansur et al., 1993). Genome-wide insertion/deletion (InDel) markers have been used for fine mapping of important economical traits in rice (Feng et al., 2005; Li et al., 2017), wheat (Chen et al., 2005; Shang, 2009), and tomato (Su et al., 2019). In this study, a mixed major gene plus polygene inheritance model was used to analyze FW resistance in cucumber, and the major effect QTL of FW was investigated by BSA. We combined transcriptome data and mapping data to detect the key candidate resistance genes in cucumber after inoculation with Foc.

MATERIALS AND METHODS

Cucumber Materials and Treatments

The two cucumber inbred lines, Superina and Rijiecheng, that showed susceptibility and resistance to FW, respectively, were used in this study. Superina (female parent, P_1) and Rijiecheng (male parent, P_2) were used to construct different generations, namely, F_1, B_1 ($F_1 \times$ Superina), B_2 ($F_1 \times$ Rijiecheng), and F_2. The F_1 generation ($P_1 \times P_2$) was planted in Autumn 2016 in a greenhouse at the experimental farm of the Department of Horticulture in Yangzhou University, and the B_1, B_2, and F_2 generations were obtained by backcrossing the F_1 plants with Superina (B_1) or Rijiecheng (B_2) and self-crossing the F_1 plants (F_2) in Spring 2017. The 200 P_1, 200 P_2, 200 F_1, 200 B_1, 200 B_2,

and 1500 F_2 plants were planted in 36-cavity plates filled with aseptic organic substrates (N, P, and K = 40–60 g/kg, humus ≥350 g/kg, pH = 6.5–7.5) in the greenhouse in Autumn 2017. Another 500 F_2 and 200 F_2 were planted in Autumn 2018 and Spring 2019, respectively. Seedlings at the second true leaf stage were inoculated with the Foc suspension (concentration 10^6 conidia/ml) using the root irrigate method. The disease classification scale was as follows: grade 0, asymptomatic; grade 1, slightly discolored stem base and cotyledon; grade 2, necrotic patches at the stem base and slight wilting of the seedling; grade 3, color of necrotic patches at the stem was deep with longitudinal fissure and visible wilting of the seedling; and grade 4, the seedling was dead (modified from Zhou et al., 2010). The disease index was calculated as follows:

$$Dicrease\ index\ (\%) = \frac{\sum (Grade \times Corresponding\ number\ of\ pathogenic\ seedlings)}{Total\ number\ of\ seedlings\ investigated\ x\ Highest\ disease\ grade} x100$$

Phenotypic Statistics and Experimental Design

Ten days after Foc inoculation, the disease grade of the seedlings was recorded and analyzed. The extremely resistant and extremely sensitive seedlings were separately pooled into an R-pool and S-pool, respectively. Genomic regions that showed signatures of resistance to FW were detected by whole genome resequencing of the DNA from the two parents and the two pools. Regions that underwent specific selection in the opposite directions were selected (Hiroki et al., 2013). The reads in the R- and S-pools were mapped to the Cucumber (Chinese Long) v2 Genome (http://cucurbitgenomics.org/organism/2). The single nucleotide polymorphisms (SNP)-index and Euclidean distance of the two pools were calculated and compared. The Euclidean distance was estimated as described by Hill et al. (2013). The SNP-indexes of the two pools were different because of genotype selection and knock-on effects. QTLs related to the FW resistance trait were roughly located by taking the intersection of the above two results (SNP-index and Euclidean distance).

Genetic Analysis

The FW resistance in the six generations of cucumber (P_1, P_2, F_1, B_1, B_2, and F_2) was analyzed using the mixed major gene plus polygene inheritance model (Gai and Wang, 1998). The maximum log-likelihood value and Akaike information criterion (AIC) were obtained by estimating the parameters of each generation and component distributions using the iterated EMC (IEMC) algorithm (Gai and Wang, 1998). After repeated iteration, the algorithm converged to a relatively stable and consistent result. The optimal model was selected according to the AIC, and the corresponding component distribution parameters were obtained. The first- and second-order parameters of the optimal model were estimated using a least squares method.

DNA Extraction, Identification of InDel Markers, and Genotyping

Total genomic DNA was extracted from cotyledons of the parent seedlings using the Cetyltrimethyl Ammonium Bromide

(CTAB) -acidic phenol extraction method (Xu et al., 2017). The concentration and quality of the extracted DNA were determined by ultraviolet spectrophotometer (Thermo Fisher, USA) and 1% agarose gel electrophoresis. The primer pairs used for screening for InDel markers were designed using Primer Premier 5.0 with the Cucumber v2 Genome sequence as the reference. For each sample, the PCR mixture (20 μl total volume) contained 2 μl 10× buffer, 1 μl dNTPs (10 mM), 1 μl primer F (50 ng/μl), 1 μl primer R (50 ng/μl), 1 μl DNA, 0.4 μl Taq DNA polymerase (10 U/μl), and 13.6 μl diethyl pyrocarbonate water (DEPC water). The touchdown PCR amplifications were performed using an Eppendorf Mastercycler Pro (Eppendorf). Subsequently, 1 μl PCR product was detected and analyzed by polyacrylamide gel electrophoresis.

PCR and polyacrylamide gel electrophoresis were used to identify polymorphisms in the F_2 generation using the parent Superina and Rijiecheng plants, and the F_1 plants are the reference. By combining the genotype and phenotype of the F_2 seedlings, we determined the location of the interval associated with resistance on the cucumber genome sequence. We designed new InDel marker primers in the resistance interval to fine-map the interval associated with resistance and repeated this until the interval was small or until it contained no new markers.

Validation of Gene Expression by Real-Time PCR (qPCR)

We have transcriptome data for Rijiecheng, a relatively resistant line, at 0, 24, 48, 96, and 192 h after inoculation with Foc (unpublished data). We used the Foc-inoculated and water-inoculated Rijiecheng seedlings to verify the expression levels of candidate genes by qPCR. Total RNA of each sample was isolated using a Mini BEST Plant RNA Extraction Kit (TaKaRa, China) and then dissolved in Ultra Pure™ DNase/RNase-free distilled water (Invitrogen, USA). The total RNA was reverse-transcribed using a PrimeScript™ RT reagent kit with genomic DNA (gDNA) eraser (TaKaRa, China). Primer sequences were designed using Beacon Designer 7.0 and screened using SeqHunter 1.0 The qPCRs were performed using SYBR® Premix Ex Taq™ (TaKaRa, China), according to the manufacturer's instructions. SYBR Green PCR cycling was performed on an IQTM5 Multicolor Real-Time PCR Detection System (Bio-Rad, USA) using 20 μl samples as follows: 95 for 3 min, followed by 39 cycles of 95 for 10 s, 60 for 20 s, and 72 for 20 s. The relative quantization of gene expression was calculated and normalized to tubulin alpha chain (*Csa4G000580*). Three biological replicates from each condition were used for qPCRs.

RESULTS

Variations in FW Phenotypes Among the Cucumber Parents and Segregated Populations

Ten days after Foc inoculation, the disease symptom grades and disease indexes of the seedlings of different generation were recorded and calculated. The disease indexes of P_1, P_2, F_1, B_1, B_2, and F_2 were 73.21, 18.13, 51.43, 48.24, 37.68, and 51.71, respectively (**Figure 1**). The average disease index of F_1 plants was higher than the mid parent value, indicating that the resistance traits of F_1 plants tended to be from the male parent Rijiecheng. The disease indexes of the two backcross generations (B_1 and B_2) shifted towards the backcross parent. The F_2 population showed positive correlation between the distribution of disease grades and obvious quantitative genetic characteristics using three replicates (**Figure 2**), which corresponds to the genetic characteristics of a mixed major gene plus polygene inheritance model.
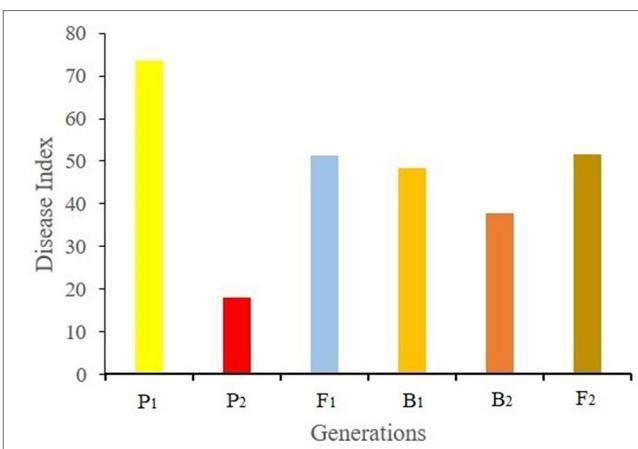
FIGURE 1 | Disease index of each generation. F_2 plants were obtained from the cross between Superina (female parent, P_1) and Rijiecheng (male parent, P_2). Superina (female parent, P_1) and Rijiecheng (male parent, P_2) were used to construct different generations, namely F_1, B_1 (F_1 × Superina), B_2 (F_1 × Rijiecheng), and F_2.
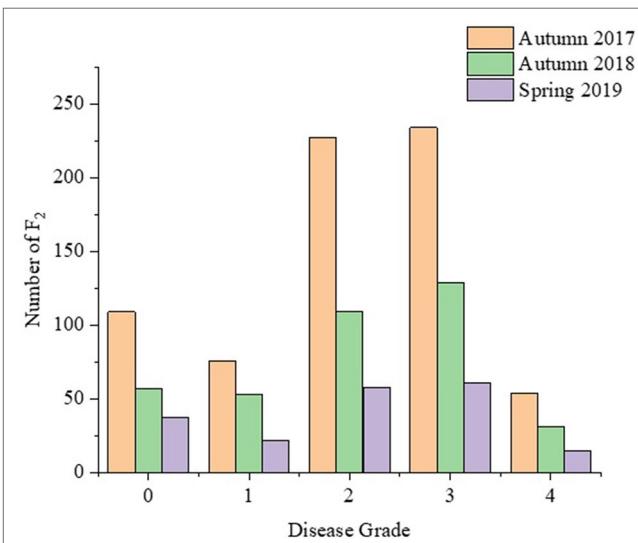
FIGURE 2 | Frequency distribution of the disease grades among the F_2 population. The disease grade was recorded 10 days after inoculation in Autumn 2017 and 2018, and Spring 2019.

Best-Fitting Genetic Model for FW Resistance and the Affecting Factors

To obtain a genetic model of cucumber FW resistance, we analyzed the segregation of the resistant phenotype among the six generations using the mixed major gene plus polygene inheritance model. We analyzed and screened a total of 18 genetic models including five categories by combining the maximum log-likelihood value and AIC. The five categories were as follows: one major gene model, two major genes model, polygene model, one major gene plus polygene model, and two major genes plus polygene model (**Table 1**). We selected the two major genes plus polygene models (E, E-1, E-2, and E-3) that had the smallest AIC values as the candidate models. Fitness tests of these four models, including equal distribution (U12, U22, and U32), Smirnov (nW2), and Kolmogorov tests (Dn) (Xu et

al., 2017), indicated the significance levels for the E, E-1, E-2, and E-3 models were 7, 1, 11, and 11, respectively (**Table 2**). Finally, we selected the E-1 model (two pairs of additive-dominance-epistatic major genes and additive-dominance polygenes) as the optimal model for resistance to FW on the basis of the combined AIC values and goodness-of-fit test.

To determine the genetic effects of major genes, we estimated first- and second-order distribution parameters for resistance to FW in the F_1 generation using the optimal E-1 model (**Table 3**). The additive effect values of the major gene and polygene were both 0.05, and the dominant effect values of the major gene and polygene were 0.99 and 0.57, respectively, which indicated that the additive effect of the major gene was consistent with that of the polygene, whereas the dominant effect of the major gene was greater than that of the polygene. This suggested that the contribution rates of major gene and polygene were consistent with inheritance of resistance to FW. The heritability of major gene (h_{mg}^2) from B_1, B_2, and F_2 generations was 22.91%, 29.70%, and 59.73%; the heritability of polygene (h_{pg}^2) from B_1, B_2, and F_2 was 29.92%, 21.78%, and 2.43%; and the environmental variance (σ_e^2) accounted for 47.62%, 48.52%5, and 37.84% of the phenotypic variance (σ_p^2) for B_1, B_2 and F_2, respectively. These results indicated that environmental factors had a large effect on cucumber resistance to FW.

Identification of the Major Effect FW QTL by BSA

A total of 214,270,191 clean reads from the transcriptomes of Superina, Rijiecheng, S-pool, and R-pool were aligned to the Cucumber v2 Genome sequence; the Q30-level was >95%,

TABLE 1 | Maximum log-likelihood values (MLVs) and Akaike information criterion (AIC) values under various genetic models estimated using the IECM algorithm.

Model	MLV	AIC	Model	MLV	AIC
A-1	−1,479.132812	2,966.265625	D	−1,466.053711	2,956.107422
A-2	−1,479.893311	2,965.786621	D-3	−1,469.131348	2,954.262695
A-3	−1,482.719971	2,971.439941	D-4	−1,464.250366	2,944.500732
B-1	−1,442.165039	2,904.330078	E	−1,433.539062	2,903.078125*
B-2	−1,455.200073	2,922.400146	E-1	−1,440.745305	2,911.490723*
B-3	−1,585.298706	3,178.597412	E-2	−1,437.410400	2,896.820801*
B-4	−1,484.792725	2,975.585449	E-3	−1,437.569580	2,893.139160*
C	−1,466.125488	2,952.250977	E-4	−1,459.107178	2,934.214355
C-1	−1.469.656250	2,953.312500	E-5	−1,459.093994	2,936.187988

A, one major gene model; *B*, two major genes model; *C*, polygene model; *D*, one major gene plus polygene model; *E*, two major genes plus polygene model. * indicated the four candidate models with smallest AIC values.

TABLE 2 | Goodness-of-fit of the two major genes plus polygene models of resistance to FW.

Model	Generation	Maximum log-likelihood estimates of component distribution parameters				
		U_1^2	U_2^2	U_3^2	$_nW^2$	D_n
E	P_1	0.03 (0.86)	0.33 (0.56)	2.73 (0.10) *	0.72 (>0.05)	0.30 (>0.05)
	P_2	0.12 (0.73)	0.01 (0.92)	2.91 (0.09) *	0.62(>0.05)	0.27 (>0.05)
	F_1	0.01 (0.94)	0.17 (0.68)	3.80 (0.05) *	0.72 (>0.05)	0.28 (>0.05)
	B_1	1.96 (0.16) *	1.58 (0.21) *	0.16 (0.69)	0.58 (>0.05)	0.28 (>0.05)
	B_2	1.23 (0.27) *	1.46 (0.23) *	0.30 (0.59)	0.55 (>0.05)	0.25 (>0.05)
E-1	P_1	0.09 (0.76)	0.11 (0.74)	0.02 (0.88)	0.60 (>0.05)	0.28 (>0.05)
	P_2	0.00 (1.00)	0.01 (0.92)	0.18 (0.67)	0.52 (>0.05)	0.24 (>0.05)
	F1	0.00 (1.00)	0.01 (0.91)	0.19 (0.66)	0.58 (>0.05)	0.25 (>0.05)
	B_1	0.01 (0.92)	0.00 (0.95)	0.45 (0.50)	0.36 (>0.05)	0.19 (>0.05)
	B_2	0.03 (0.87)	0.04 (0.85)	0.02 (0.89)	0.40 (>0.05)	0.21 (>0.05)
	F_2	0.00 (0.97)	0.08 (0.78)	1.61 (0.21) *	3.06 (>0.05)	0.16 (>0.05)
E-2	P_1	0.08 (0.78)	0.02 (0.90)	2.55 (0.11) *	0.71 (>0.05)	0.27 (>0.05)
	P_2	0.54 (0.46) *	0.07 (0.80)	3.33 (0.07) *	0.67 (>0.05)	0.29 (>0.05)
	F_1	0.00 (0.98)	0.23 (0.64)	3.98 (0.05) *	0.73 (>0.05)	0.28 (>0.05)
	B_1	8.40 (0.00) *	9.22 (0.00) *	0.85 (0.36) *	1.25 (>0.05)	0.37 (>0.05)
	B_2	1.77 (0.18) *	2.28 (0.13) *	0.80 (0.37) *	0.63 (>0.05)	0.26 (>0.05)
	F_2	0.00 (0.97)	0.20 (0.66)	2.76 (0.10) *	3.15 (>0.05)	0.16 (>0.05)
E-3	P_1	0.00 (0.97)	0.08 (0.78)	1.61 (0.21) *	3.06 (>0.05)	0.16 (>0.05)
	P_2	0.08 (0.78)	0.03 (0.87)	3.06 (0.08) *	0.62 (>0.05)	0.26 (>0.05)
	F_1	0.05 (0.83)	0.52 (0.47) *	4.27 (0.04) *	0.74 (>0.05)	0.30 (>0.05)
	B_1	7.82 (0.01) *	8.58 (0.00) *	0.79 (0.37) *	1.19 (>0.05)	0.36 (>0.05)
	B_2	1.45 (0.23) *	1.82 (0.18) *	0.53 (0.47) *	0.60 (>0.05)	0.25 (>0.05)
	F_2	0.15 (0.70)	0.00 (0.98)	2.43 (0.12) *	3.16 (> 0.05)	0.17 (>0.05)

U_1^2, U_2^2, and U_3^2, uniformity test; $_nW^2$, Smirnov test; D_n, Kolmogorov test; FW, Fusarium wilt. * indicates significance at 0.05.

TABLE 3 | Estimates of genetic parameters for resistance to FW in the F_1 plants (Superina×Rijecheng) cross under the E-1 model

1st parameter	Estimate	2nd parameter	Estimate		
			B_1	B_2	F_2
m	0.99	σ_P^2	1.26	1.23	1.58
da	0.05	σ_{mg}^2	0.29	0.37	0.94
ha	0.99	σ_{pg}^2	0.38	0.27	0.04
db	0.05	σ_e^2	0.60	0.60	0.60
hb	0.57	h_{mg}^2	22.91	29.70	59.73
ha/da	20.27	h_{pg}^2	29.92	21.78	2.43
hb/db	11.94	$1-h_{mg+pg}^2$	47.17	48.52	37.84

TABLE 4 | Evaluation of sample sequencing data.

Name	Clean Reads	Clean Base	Q30 (%)	GC (%)	Average Depth	Radio of Average Depth (%)
Superina	30,821,023	9,212,830,776	93.13	36.97	21	98.56
Rijiecheng	30,766,127	9,197,436,830	93.15	37.23	23	98.80
S-pool	76,119,559	22,745,447,354	92.91	37.30	58	99.10
R-pool	76,572,482	22,865,387,058	92.90	37.24	60	99.12

and the GC-contents was >35%. The detailed statistics for each transcriptome including numbers of clean reads, numbers of bases, and average depth are given in **Table 4**. We identified a 3.78 Mb region on chromosome 2 (Chr2) that contained 625 genes that had an association threshold (Euclidean distance) of 0.12. One region on Chr2 with a total length of 1.91 Mb and containing 319 genes had an association threshold of 0.25 after fitting the △SNP-index. We used the intersection of these two results to identify the region of the genome associated FW resistance. The identified region was on Chr2, was 1.91 Mb long, and contained 319 genes (**Figure 3**), and was designated as the major effect QTL (*fw2.1*) related to the FW resistance trait.

Fine Mapping QTL *fw2.1* to 0.6 Mb on Chr2
To narrow down the major effect QTL identified by BSA, whole genome resequencing of the two parents was performed to confirm the high-quality InDels. We genotyped 500 F_2 plants using two

polymorphic InDel markers (InDel89749 and InDel1817308). Three other InDel markers that were evenly distributed in the major effect QTL interval were designed. Six different genotypes were detected among the F_2 plants; however, two of them had disease grades 0 or 2 (**Figure 4**). By combining the genotypes and phenotypes of the F_2 seedlings, the candidate interval was located between InDel1248093 and InDel1817308, which are at a physical distance of 569,215 bp in a region that contains 80 genes. Thus, the major effect QTL was fine-mapped to an approximately 0.60 Mb interval (from 1,248,093 to 1,817,308 bp) on Chr2 of the Cucumber v2 Genome assembly.

Analysis of Candidate Genes by RNA-Seq
To confirm the FW resistance genes, we analyzed the 80 candidate genes in the major effect QTL using RNA-sequencing (RNA-seq). We combined the RNA-seq data with the previously obtained transcriptome data [national center for biotechnology information's (NCBI's) sequence readarchive (SRA) database: PRJNA472169] to detect differentially expressed genes (DEGs) in the 0.60 Mb interval on Chr2. All the DEGs were positively regulated, except for *Csa2G008760*. DEGs were defined as having a false discovery rate ≤0.01 and fold change ≥2 or ≤−2. Twelve genes were found to be differentially induced by FW (**Figure 5**), namely, *Csa2G007990* (calmodulin), *Csa2G008030* [probable Guanosine diphosphate (GDP)-mannose transporter 2], *Csa2G008110* (monogalactosyldiacylglycerol synthase), *Csa2G008760* (chitinase), *Csa2G008770* (adenylate kinase), *Csa2G008780* and *Csa2G009330*
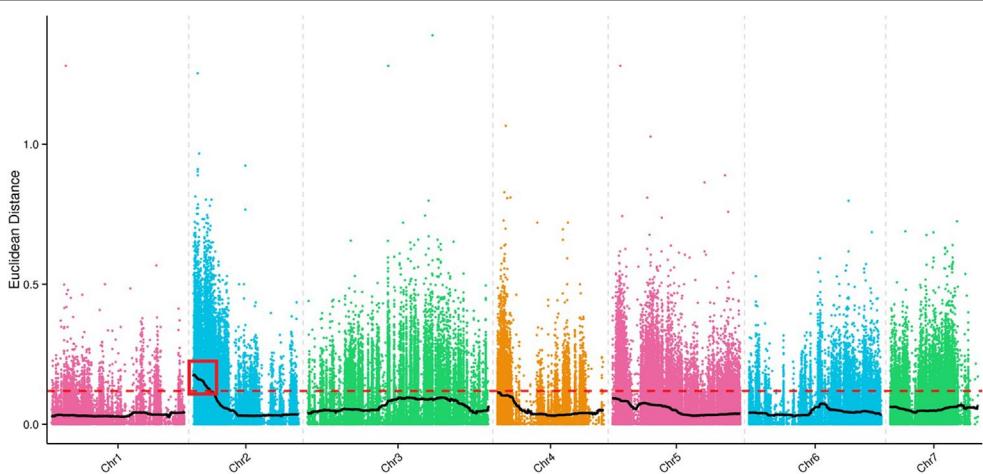
FIGURE 3 | Identification of the major effect *Fusarium* wilt (FW) quantitative trait locus (QTL). Colored points indicate the Euclidean distance values of every SNP on each of the chromosomes. The black lines indicate the Euclidean distance values after matching. The red dotted line is the threshold line. Values above the threshold line were used to select the candidate FW trait-related interval as shown in the red box.
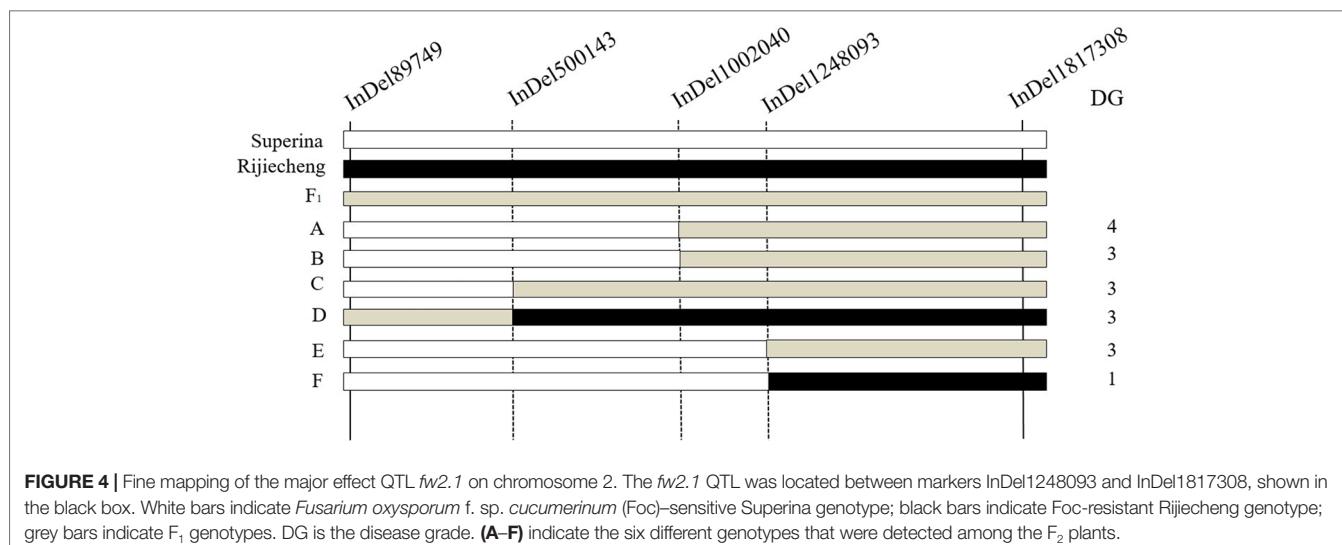
FIGURE 4 | Fine mapping of the major effect QTL *fw2.1* on chromosome 2. The *fw2.1* QTL was located between markers InDel1248093 and InDel1817308, shown in the black box. White bars indicate *Fusarium oxysporum* f. sp. *cucumerinum* (Foc)–sensitive Superina genotype; black bars indicate Foc-resistant Rijiecheng genotype; grey bars indicate F₁ genotypes. DG is the disease grade. **(A–F)** indicate the six different genotypes that were detected among the F₂ plants.

(unknown proteins), *Csa2G009300* (DNA replication licensing factor MCM5), *Csa2G009360* (RING-type finger protein 126), *Csa2G009430* (transmembrane protein), *Csa2G009440* (serine-rich protein), and *Csa2G009470* (betaine aldehyde dehydrogenase). We also analyzed the expression patterns of the 12 candidate genes in Rijiecheng by qPCR (**Figure 6**; **Supplementary Figure 1**). The primers of these 12 genes are shown in **Supplementary Table 1**. We found that 5 of the 12 genes were significantly induced by FW, and their relative expression levels were higher in Foc-inoculated Rijiecheng than in the water-inoculated plants.

DISCUSSION

Several studies on the inheritance of resistance in cucumber investigated quantitative resistance traits controlled by multiple

FIGURE 5 | Transcriptional profiles of 12 candidate genes in the major effect QTL *fw2.1* on chromosome 2 at different times after inoculation. The expression data were converted to Log₂ fragments per kilobase million (FPKM) to measure the expression levels of the FW resistance genes. The color scale indicates fold changes in the expression levels of the genes induced by FW.

genes (Pierce and Wehner, 1990; Liu et al., 2003; Wang et al., 2005) and qualitative resistance traits controlled by a single gene (Netzer et al., 1977; Vakalounakis 2015). Contradictory results were obtained, possibly because there are different physiological races of Foc with different pathogenicities in different countries (Weng et al., 1989; Armsrong et al., 1987). We investigated the inheritance of cucumber FW resistance using Rijiecheng and Superina as the parents and concluded that the resistance in Rijiecheng was quantitative and the inheritance of FW resistance was controlled by multiple genes. This result was confirmed for F₂ plants grown in different years and seasons, indicating that resistance to FW was a quantitative trait in cucumber. In dry pea, an F₂ population segregating for high levels of resistance to *Fusarium solani* f. sp. Pisi also had a disease reaction phenotype in repeated greenhouse trials (Coyne et al., 2019).

Many studies on molecular linkage markers and genetic mapping of cucumber powdery mildew and downy mildew have been reported (Zhang et al., 2011; Zhang et al., 2013a; Zhang et al., 2013b; Xu et al., 2016). However, the genetic mapping of cucumber FW has rarely been reported (Zhang et al., 2014). One simple sequence repeat (SSR) marker linked to cucumber *Foc2.1* was identified in a genetic interval of 5.98 cM (Wang, 2005) and validated among 46 germplasms. The accuracy rate of this SSR marker for selecting resistant germplasm was 87.88%. Zhang et al. (2014) found one major QTL on cucumber Chr2 (SSR03084–SSR17631) for FW resistance in a genetic distance of 2.4 cM. Zhou et al. (2015) concluded that cucumber *Foc4* resistance to FW was located between SSR17631 and SSR00684 on Chr2. In this study, we found five InDel markers with polymorphisms in P₁, P₂, and F₁ at different physical positions on Chr2 by BSA and successfully identified the major effect QTL on Chr2 with a physical distance of 0.60 Mb (InDel1248093–InDel1817308). Notably, the physical positions of all these QTLs are different: 2,526,888–3,262,528 (Zhang et al., 2014), 3,255,681–3,262,528 (Zhou et al., 2015), and 1,248,093–1,817,308 in the present study.

We found 12 DEGs in *fw2.1* by combining BSA with RNA-seq after inoculating the plants with Foc infection and taking samples at
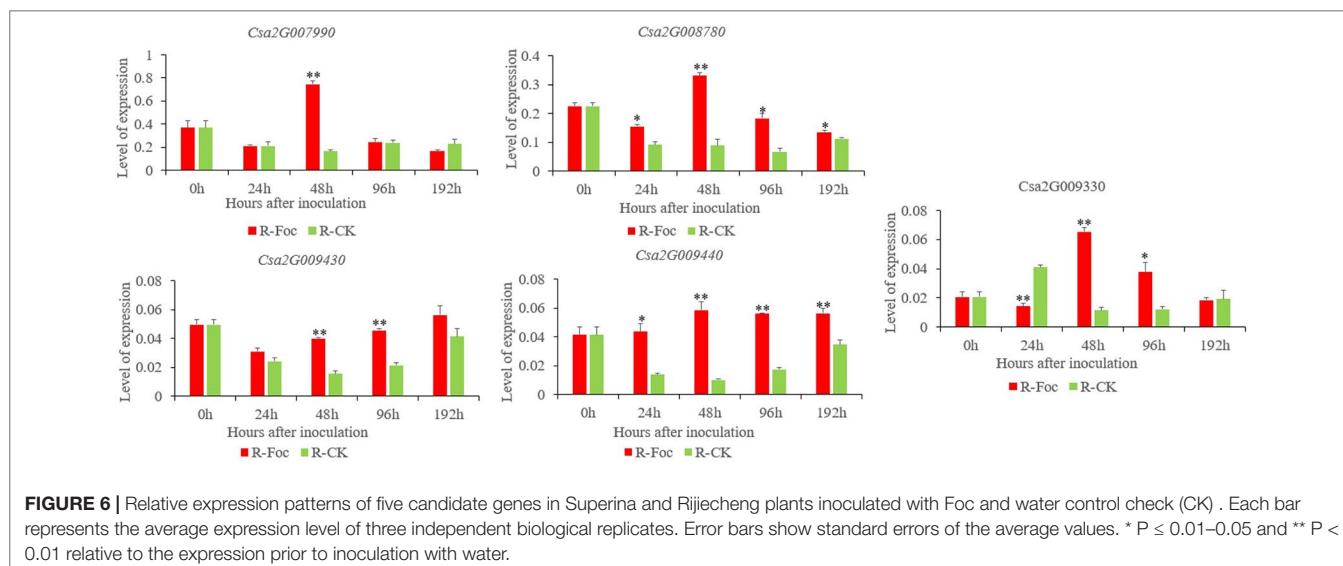
FIGURE 6 | Relative expression patterns of five candidate genes in Superina and Rijiecheng plants inoculated with Foc and water control check (CK) . Each bar represents the average expression level of three independent biological replicates. Error bars show standard errors of the average values. * P ≤ 0.01–0.05 and ** P < 0.01 relative to the expression prior to inoculation with water.

different times. In previous studies, RNA-seq data were used to explore key susceptible genes in cucumber that responded to foliage diseases (Zheng et al., 2019). Liu et al. (2019) suggested that the differences in gene expression following cucumber mosaic virus (CMW) infection might explain the different resistance levels of two lines on the basis of RNA-seq data. Shi et al. (2018) analyzed the RNA-seq data of a recombinant inbred line and located a QTL on Chr5 that contained nine genes in cucumber infected with CMW. We found five candidate genes by combining QTL mapping and RNA-seq for resistance to FW in cucumber. *Csa2G007990* encodes calmodulin and is highly expressed in cucumber. Naveed et al. (2019) found that when calmodulin bound to the effector peptide protein Avrblb2, it incited a hypersensitive response. *Csa2G009430* encodes a transmembrane protein that has been widely studied in animal disease resistance, but not in plants. *Csa2G009440* encodes a serine-rich protein, which has been linked to the ability of bacteria to attach to the hosts. For example, Santiago et al. (2007) suggested that serine-rich region may be involved in the negative regulation of phytochrome signal transduction, which is known to yield a hyperactive photoreceptor. *Csa2G008780* and *Csa2G009330* are highly expressed in cucumber, and their expression levels were significantly higher in Foc-inoculated plants than that in water-inoculated plants. The functions of these two genes are still unknown and require further study to determine their characteristics and functions. We plan to verify the functions of the five candidate genes using other methods such as virus-induced gene silencing and overexpression. Our results will help to better understand the genetic mechanisms and provide a strong basis for fine mapping of the major effect QTL and for cloning the candidate genes for resistance to FW in cucumber.

DATA AVAILABILITY STATEMENT

The transcriptome data associated with this study can be found in NCBI using accession number PRJNA472169 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA472169).

AUTHOR CONTRIBUTIONS

XC and XX conceived the experiment. JD and JX performed the research and wrote the manuscript. JD and QX analyzed the data. All authors reviewed and approved this submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01425/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Relative expression patterns of the seven rejected candidate genes in the Superina and Rijiecheng plants inoculated with Foc and water (CK). Each bar represents the average expression level of three independent biological replicates. Error bars show standard errors of the average values. *P ≤0.01–0.05 and **P <0.01 relative to the expression prior to inoculation by water.

REFERENCES

Armsrong, G. M., Armstrong, J. K., and Neter, U. (1987). Pathogenic races of the cucumber-wilt Fusarium. *Plant Dis. Rep.* 62, 824–828.

Cao, Q. H., Wan, J., and Chen, J. F. (2007). Research advances on downy mildew resistance in cucumber. *China Cucurbits Veg* 1, 27–30. doi: 10.16861/j.cnki.zggc.2007.01.013

Chen, J. B., Jing, R. L., Yuan, H. Y., Wei, B., and Chang, X. P. (2005). Single nucleotide polymorphism of TaDREB1 gene in Wheat germplasm. *Sci Agric Sin.* 38, 2387–2394. doi: 10.1109/CIMCA.2005.1631314

Coyne, C. J., Porter, L. D., Boutet, G., Ma, Y., Mcgee, R. J., Lesne, A., et al. (2019). Confirmation of Fusarium root rot resistance QTL Fsp-Ps 2.1 of pea under controlled conditions. *BMC Plant Biol.* 98, 1–8. doi: 10.1186/s12870-019-1699-9

Falconer, D. S., and Mackay, T. F. C. (1996). Introduction to Quantitative Genetics, 4th ed. *Am. J. Hum. Genet.* 46, 1231. doi: 10.1016/0168-9525(96)81458-2

Feng, F. J., Luo, L. J., Li, Y., Zhou, L. G., Xu, X. Y., Wu, X. Y., et al. (2005). Comparative analysis of polymorphism of InDel and SSR markers in Rice. *Mol. Plant Breed.* 3, 725–730. doi: 10.1080/00207177508921984

Gai, J. Y., and Wang, J. K. (1998). Identification and estimation of a QTL model and its effects. *Theor. Appl. Genet.* 97, 1162–1168. doi: 10.1007/s001220051005

Giovannoni, J. J., Wing, R. A., Ganal, M. W., and Tanksley, S. D. (1991). Isolation of molecular markers from specific chromosome intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 19, 6553–6568. doi: 10.1093/nar/19.23.6553

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., Su, Y. C., and Yost, H. J. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23, 687–697. doi: 10.1101/gr.146936.112

Hiroki, T., Akira, A., Kentaro, Y., Shunichi, K., Satoshi, N., Chikako, M., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Li, G., Carver, B. F., Cowqer, C., Bai, G., and Xu, X. (2018). Pm223899, a new recessive powdery mildew resistance gene identified in Afghanistan landrace PI 223899. *Theor. Appl. Genet.* 131 (12), 2775–2783. doi: 10.1007/s00122-018-3199-y

Li, Z. Z., Xu, Q. K., Yu, H. P., Zhou, T. T., Xue, D. W., Zeng, D. L. et al. (2017). Genetic analysis and gene mapping of yellow leaf and dwarf (yld) mutant in rice. *Acta Agronomica Sinica.* 4, 522–529. doi: 10.3724/SP.J.1006.2017.00522

Liu, D. L., Yang, R. H., and Ha, Y. H. (2003). Research on genetic characteristics of cucumbers resisting to Fusarium wilt. *Tianjin Agric. Sci.* 9, 33–35. doi: 10.3969/j.issn.1006-6500.2003.02.011.

Liu, D., Cheng, Y. Z., Gong, M., Zhao, Q., Jiang, C. H., Cheng, L. R., et al. (2019). Comparative transcriptome analysis reveals differential gene expression in resistant and susceptible tobacco cultivars in response to infection by cucumber mosaic virus. *Crop J.* 7, 307–321. doi: 10.1016/j.cj.2018.11.008

Mansur, L. M., Orf, J., and Lark, K. G. (1993). Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (Glycine max L. Merr.). *Theor. Appl. Genet.* 86, 914–918. doi: 10.1007/BF00211041

Martínez, R., Aguilar, M. I., Guirado, M. L., Álvarez, A., and Gómez, J. (2003). First report of fusarium wilt of cucumber caused by Fusarium oxysporum in Spain. *Plant Pathol.* 52, 410. doi: 10.1046/j.1365-3059.2003.00832.x

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828

Naveed, Z. A., Bibi, S., and Ali, G. S. (2019). The Phytophthora RXLR effector Avrblb2 modulates plant immunity by interfering with Ca2+ signaling pathway. *Front Plant Sci.* 10. doi: 10.3389/fpls.2019.00374

Netzer, D., Niegro, S., and Galun, E. (1977). A dominant gene conferring resistance to Fusarium wilt in cucumber. *Phytopathology* 77, 525–527. doi: 10.1094/Phyto-67-525

Owen, J., H. (1955). Fusarium wilt of cucumber. *Phytopathology* 45, 435–443. doi: 10.1111/j.1399-3054.1955.tb07793.x

Pierce, L. K., and Wehner, T. C. (1990). Review of genes and linkage groups in cucumber. *Hort Sci.* 25, 605–615. doi: 10.21273/HORTSCI.25.6.605

Pu, Z. J., Zhang, Y. J., Liu, D., Dai, L. T., and Wang, W. B. (2011). Research progress in biological control strategies for Fusarium wilt of cucumber. *China Vegetables* 6, 9–14.

Santiago, A. T., Dimitry, D., Andreas, H., Christian, F., and Jorge, J. C. (2007). The serine-rich N-terminal region of *Arabidopsis* phytochrome A is required for protein stability. *Plant Mol. Biol.* 63, 669–678. doi: 10.1007/s11103-006-9115-x

Shang, S. J. D. (2009). Development and application of RGA markers in wheat. *Chin. Acad. Agric. Sci.*

Shi, L. X., Yang, Y. H., Xie, Q., Miao, H., Bo, K. L., Song, Z. C., et al. (2018). Inheritance and QTL mapping of cucumber mosaic virus resistance in cucumber (Cucumis Sativus L.). *PloS One* 13, 1–12. doi: 10.1371/journal.pone.0200571

Su, X., Zhu, G., Huang, Z., Wang, X., Guo, Y., Li, B., et al. (2019). Fine mapping and molecular marker development of the Sm gene conferring resistance to gray leaf spot (Stemphylium spp.) in tomato. *Theor. Appl. Genet.* 132 (4), 871–882. doi: 10.1007/s00122-018-3242-z

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Vakalounakis, D. J. (2015). Inheritance and linkage of resistance in cucumber line SMR-18 to races 1 and 2 of Fusarium oxysporum f. sp. cucumerium. *Plant Pathol.* 44, 169–172.

Vakalounakis, D. J. (1993). Inheritance and genetic linkage of fusarium wilt (*Fusarium oxysporum* f. sp. cucumerinum race 1) and scab (Cladosporium cucumerinum) resistance genes in cucumber (Cucumis sativus L.). *Ann. Appl. Biol.* 123, 359–365. doi: 10.1111/j.1744-7348.1993.tb04098.x

Wang, Y. J. (2005). *Studies on molecular marker of fusarium wilt resistance related gene in cucumber (Cucumis sativus L.).* Northwest A&F University. 1–55.

Weng, Z. X., Xu, X. B., and Feng, D. X. (1989). Study on the physiological race of *Fusarium oxysporum* f. sp. cucumerinum Owen. *Chin. Vegetables* 1, 19–21.

Xu, X. W., Ji, J., Xu, Q., Qi, X., and Chen, X. (2017). Inheritance and quantitative trail loci mapping of adventitious root numbers in cucumber seedlings under waterlogging conditions. *Mol. Genet. Genomics* 292, 353–364. doi: 10.1007/s00438-016-1280-2

Xu, X., Yu, T., Xu, R., Shi, Y., Lin, X. Y., Xu, Q., et al. (2016). Fine mapping of a dominantly inherited powdery mildew resistance major-effect QTL, Pm1.1, in cucumber identifies a 41.1kb region containing two tandemly arrayed cysteine-rich receptor-like protein kinase genes. *Theor. Appl. Genet.* 129 (3), 507–516. doi: 10.1007/s00122-015-2644-4

Ye, S. F., Yu, J. Q., Peng, Y. H., Zheng, J. H., and Zou, L. Y. (2004). Incidence of *Fusarium* wilt in *Cucumis sativus* L. is promoted by cinnamic acid, an autotoxin in root exudates. *Plant Soil* 263, 143–150. doi: 10.1023/B:PLSO.0000047721.78555.dc

Yuan, F., Zhang, C. L., and Shen, Q. R. (2003). Alleviating effect of phenol compounds on cucumber *fusarium* wilt and mechanism. *Agric. Sci. China* 2, 647–652.

Zhang, K., Wang, X., Zhu, W., Qin, X., Xu, J., Cheng, C., et al. (2018). Complete resistance to powdery mildew and partial resistance to downy mildew in a Cucumis hystrix introgression line of cucumber were controlled by a co-localized locus. *Theor. Appl. Genet.* 131 (10), 2229–2243. doi: 10.1007/s00122-018-3150-2

Zhang, S. P., Liu, M. M., Miao, H., Zhang, S. Q., Yang, Y. H., Xie, B. Y., et al. (2011). QTL mapping of resistance genes to powdery mildew in cucumber (Cucumis sativus L.). *Sci Agric Sin.* 129 (3), 507–516. doi: 10.3864/j.issn.0578-1752.2011.17.011

Zhang, S. P., Liu, M. M., Miao, H., Zhang, S. Q., Yang, Y. H., Xie, B. Y., et al. (2013a). Chromosomal mapping and QTL analysis of resistance to downy mildew in Cucumis sativus. *Plant Dis.* 97, 245–251. doi: 10.1094/PDIS-11-11-0941-RE

Zhang, S. P., Miao, H., Yang, Y. H., Xie, B. Y., Wang, Y., and Gu, X. F. (2014). A major quantitative trait locus conferring resistance to fusarium wilt was detected in cucumber by using recombinant inbred lines. *Mol. Breed.* 34, 1805–1815. doi: 10.1007/s11032-014-0140-1

Zhang, S., Miao, H., Sun, R., Wang, X., Huang, S. W., Wehner, T. C., et al. (2013b). Localization of a new gene for bitterness in cucumber. *J. heredity* 104, 134–139. doi: 10.1093/jhered/ess075

Zheng, X. Y., Yang, J. J., Lou, T. X., Zhang, J., Yu, W. J., and Wen, C. L. (2019). Transcriptome profile analysis reveals that CsTCP14 induces susceptibility to Foliage Disease in cucumber. *Int. J. Mol. Sci.* 20, 2582. doi: 10.3390/ijms20102582

Zhou, H. M., Mao, A. J., Zhang, L. R., Zhang, F., Wang, Y. J., and Yang, W. C. (2010). Research on inoculation method and the inheritance of resistance to fusarium oxysporum f. sp. cucumerinum on cucumber. *Acta Agric. Boreali-Sinica* 25, 186–190. doi: 10.3724/SP.J.1238.2010.00569

Zhou, X. G., Wu, F. Z., Wang, X. Z., and Yuan, Y. (2008). Progresses in the mechanism of resistance to fusarium wilt in cucumber. *J. Northeast Agric. Univ.* 15, 1–6.

Zhou, H. M., Dong, C. J., Zhang, H. Y., Ren, Y., Guo, S. G., Yang, W. C., et al. (2015). SSR molecular markers and location of the gene Foc-4 linked to the resistance to Fusarium Wilt of cucumber. *Mol. Plant Breed.* 13, 1980–1986. doi: 10.13417/j.gab.034.001980

Zhou, X. G., and Wu, F. Z. (2012). Dynamics of the diversity of fungal and *Fusarium* communities during continuous cropping of cucumber in the greenhouse. *FEMS Microbiol. Ecol.* 80, 469–478. doi: 10.1111/j.1574-6941.2012.01312.x

Mapping Cucumber Vein Yellowing Virus Resistance in Cucumber (*Cucumis sativus* L.) by Using BSA-seq Analysis

Marta Pujol[1,2], Konstantinos G. Alexiou[1,2], Anne-Sophie Fontaine[3], Patricia Mayor[4], Manuel Miras[4], Torben Jahrmann[3], Jordi Garcia-Mas[1,2] and Miguel A. Aranda[4]*

[1] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Plant and Animal Genomics Program, Barcelona, Spain, [2] Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Genomics and Biotecnology Program, Barcelona, Spain, [3] Semillas Fitó S.A., Biotechnology Department, Barcelona, Spain, [4] Centro de Edafología y Biología Aplicada del Segura (CEBAS)-CSIC, Departamento de Biología del Estrés y Patología Vegetal, Murcia, Spain

Cucumber vein yellowing virus (CVYV) causes severe yield losses in cucurbit crops across Mediterranean countries. The control of this virus is based on cultural practices to prevent the presence of its vector (*Bemisia tabaci*) and breeding for natural resistance, which requires the identification of the loci involved and the development of molecular markers for linkage analysis. In this work, we mapped a monogenic locus for resistance to CVYV in cucumber by using a Bulked Segregant Analysis (BSA) strategy coupled with whole-genome resequencing. We phenotyped 135 F_3 families from a segregating population between a susceptible pickling cucumber and a resistant Long Dutch type cucumber for CVYV resistance. Phenotypic analysis determined the monogenic and incomplete dominance inheritance of the resistance. We named the locus *CsCvy-1*. For mapping this locus, 15 resistant and 15 susceptible homozygous F_2 individuals were selected for whole genome resequencing. By using a customized bioinformatics pipeline, we identified a unique region in chromosome 5 associated to resistance to CVYV, explaining more than 80% of the variability. The resequencing data provided us with additional SNP markers to decrease the interval of *CsCvy-1* to 625 kb, containing 24 annotated genes. Markers flanking *CsCvy-1* in a 5.3 cM interval were developed for marker-assisted selection (MAS) in breeding programs and will be useful for the identification of the target gene in future studies.

Keywords: Cucumber vein yellowing virus, cucumber, resistance, mapping, BSA-seq, breeding, marker-assisted selection

INTRODUCTION

Cucumber vein yellowing virus (CVYV) is an ipomovirus (family *Potyviridae*) that is transmitted in a semi-persistent manner by the whitefly *Bemisia tabaci*. CVYV was first reported in Eastern countries of the Mediterranean basin (Israel, Jordan, Turkey, and Cyprus) later expanding to Western Mediterranean countries including Spain, Portugal, France, and Tunisia (reviewed in Navas-Castillo et al., 2014). CVYV infects cucurbits, causing symptoms of variable intensity. In melon (*Cucumis melo* L.) and cucumber (*Cucumis sativus* L.), it causes a typical severe vein clearing often followed by generalized chlorosis and necrosis. In cucurbit-producing areas of heavy *B. tabaci* infestation, it can

cause epidemics with massive yield losses and dramatic economic consequences. Although significant diversity has been reported for this virus, epidemics in Western Mediterranean countries seem to be associated to genetically uniform virus populations (Desbiez et al., 2019), perhaps as a consequence of single virus introductions followed by rapid epidemic expansions (Janssen et al., 2007). At the start of the CVYV epidemics, disease control relied heavily on early detection (Martínez-García et al., 2004) and eradication, and whitefly control. Sources of resistance were soon identified in cucumber (e.g. Picó et al., 2003) and indeed, commercial seed companies are currently selling cucumber hybrids resistant to CVYV, which represent an excellent solution for disease control. Resistant accessions, varieties and hybrids seem to share the common characteristic that resistance is partial; symptoms in inoculated plants are mild or absent, and the virus can be detected infecting systemically the so-called resistant plants, although at reduced levels as compared to susceptible controls (e.g. Galipienso et al., 2013).

Breeding for CVYV resistance of cucumber varieties requires the development of molecular markers linked to the trait of interest, as pathology tests are labor-intensive and time-consuming. For molecular breeding, several genetic and genomics resources are available for cucumber, which have substantially increased after the release of reference genomes (Huang et al., 2009; Wóycicki et al., 2011; Yang et al., 2012). Cucumber has a relatively small genome (367 Mb, 2n = 2x = 14) with a very narrow genetic base (Staub et al., 2008; Wóycicki et al., 2011). The availability of high-density consensus maps and whole genome sequencing has facilitated the identification of NB-LRR resistance genes in the cucumber genome and the map-based cloning of candidate genes (Ren et al., 2009; Yang et al., 2013). These achievements are the basis for efficient marker-assisted selection (MAS). In this sense, Bulked Segregant Analysis (BSA) was developed as a rapid method for the detection of molecular markers linked to target traits in mapping populations (Michelmore et al., 1991). The principle of BSA is the selection of a small group of individuals from a segregating population that belong to phenotypic contrasting extremes of the target trait. These individuals are then pooled in two bulks, and fingerprinted to obtain genetic polymorphisms. When the trait is monogenic, the number of individuals per bulk can be reduced to 10-20, but in the case of quantitative trait loci (QTL) this number should be increased (Sun et al., 2010; Takagi et al., 2013; Zou et al., 2016). With the improvement of technologies and the significant reduction of next generation sequencing (NGS) costs, whole-genome resequencing has been coupled to BSA. The combination of BSA with NGS (BSA-seq) has accelerated the identification of tightly linked markers for important traits, improving the resolution of maps for gene identification and QTL mapping (Zou et al., 2016). In cucumber, BSA-seq has been successfully applied for mapping traits such as early flowering (Lu et al., 2014), flesh thickness (Xu et al., 2015) and downy mildew resistance (Win et al., 2017).

The aims of this work were to study the inheritance of the resistance conferred by the resistant accession CE0749 in a segregating $F_{2:3}$ population, to map the *CsCvy-1* locus by using a BSA-seq approach, and to develop molecular markers that are easily transferable to cucumber breeding programs.

MATERIALS AND METHODS

Plant Material and Phenotyping for CVYV Resistance

Accession CE0754 (hereafter P_S), a CVYV-susceptible pickling cucumber, was crossed with accession CE0749 (hereafter P_R), a CVYV-resistant Long Dutch type cucumber, to obtain the F_1, F_2 and $F_{2:3}$ segregating populations used to perform the genetic mapping of the resistance trait. P_S, P_R, F_1, F_2, and $F_{2:3}$ plants were inoculated mechanically with CVYV-AILM (Martínez-García et al., 2004) by rubbing recently-expanded cotyledons with extracts from CVYV-infected cucumber plants (cv. SMR-58) and were re-inoculated three days after. To measure virus accumulation in P_S, P_R and F_1-inoculated plants, we followed procedures described by Marco et al. (2003); quantitative dot-blot hybridization was done using the probe described by (Martínez-García et al., 2004). Plants were sampled at 9, 16, 23, and 30 days post inoculation (dpi) taking three leaf discs measuring 8 mm in diameter per leaf sampled. Symptoms were scored using a 0–3 scale: (0) No symptoms; (1) mild chlorotic mottling in young but fully-expanded leaves in interveinal petiole-proximal leaf areas; (2) similar to (1) plus vein yellowing evident in fully-expanded leaves and incipient in young developing leaves; (3) obvious vein yellowing in all leaves, including young developing leaves, chlorotic mosaics in fully-expanded leaves and overall plant growth reduction. A minimum of nine plants were used per treatment. Plants were kept in an insect-proof glasshouse, with temperature control set at 26°C/18°C (day/night) throughout the experiments.

DNA Extraction and NGS Sequencing

Young leaves from the parental lines and the F_2 population were collected, frozen in liquid nitrogen, and stored at −80°C. DNA was extracted following the CTAB method (Doyle, 1991), adding a purification step using Phenol : Chloroform:Isoamyl alcohol (25:24:1). The integrity of DNA was evaluated by agarose gel electrophoresis and quantified with the PicoGreen® dsDNA Assay Kit (Life Technologies) according to the manufacturer's protocol. For NGS sequencing, we pooled equimolar concentrations of DNA from 15 homozygous resistant F_2 plants (R-Bulk), and from 15 homozygous susceptible F_2 plants (S-Bulk). Twenty μg aliquots of each bulked DNA and both parental lines were sent to the National Centre for Genomic Analysis (CNAG-CRG, Barcelona, Spain) for library construction and sequencing. Libraries of 300 bp and 500 bp average insert size for bulks and parents, respectively, were sequenced with Illumina HiSeq 2000 (Illumina, Inc. San Diego, CA, USA), generating 2 × 100 bp paired-end reads for both datasets.

Conventional Linkage Mapping

A set of 172 polymorphic cucumber SNP markers, distributed across the seven chromosomes of the cucumber genome, were selected between P_R and P_S from the resequencing data (**Supplementary Table 1**). Kompetitive Allele Specific PCR (KASP, https://www.lgcgroup.com) was used for genotyping a subset of 72 individuals, with the 172 SNPs converted to KASP markers, following the

protocol of LGC Genomics. A genetic map was constructed using JoinMap® 5 (Kyazma, B.V.), with 172 SNP markers data and the phenotypic data used as another marker (*CsCvy-1*) due to the monogenic inheritance of the trait. Three F_2 individuals and one SNP marker were excluded because of the high amount of missing data. Two more SNP markers were excluded for being identical (similarity value = 1.000) to other nearby markers. Linkage analysis and marker order were performed with the regression mapping algorithm, and genetic distance was calculated using the Kosambi mapping function (Crow, 1990). QTL analysis was performed using MapQTL6® (Kyazma B.V.) using both interval mapping and Kruskal-Wallis (KW) analysis.

Variant Detection and BSA-seq Analysis

Variant Detection and Functional Effect Annotation

In order to detect variants of the genome linked to the resistance against CVYV, a BSA-seq strategy was implemented in which two pools of F_2 individuals were chosen depending on their phenotype. 15 homozygous resistant F_2 individuals (R-bulk) and 15 homozygous susceptible F_2 individuals (S-bulk) were selected for pooling and sequencing. Paired-end Illumina sequencing data from parental lines and the two bulks were trimmed (length ≥35 bp, with a mean sliding window of 4 bp phred quality score ≥20) using Trimmomatic (Bolger et al., 2014) and the output was quality checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed data were aligned versus the ChineseLong 9930 v3 assembly ftp://cucurbitgenomics.org/pub/cucurbit/genome/cucumber/Chinese_long/v3/) using the BWA-MEM algorithm (v0.7.16a-r1181; http://bio-bwa.sourceforge.net/bwa.shtml) with default parameters. After removal of unmapped reads and marking of PCR duplicates, variant calling was performed with samtools (v1.5; Li, 2011) using default parameters, except for the following: mapping quality ≥10 and base quality ≥20. Variant calling format (VCF) files were filtered by applying the following criteria: genotype quality ≥10, depth ≥10, biallelic sites, alternative allele frequency ≤ 0.9, no missing data.

Structural variant (SV) analysis between the two parents was conducted using DELLY (Rausch et al., 2012) and Pindel (Ye et al., 2009). Both programs were run with default parameters. Raw data underwent technical- and visual-based filtering. The technical filters applied were the following: read depth ≥10 in at least one sample, parents variable between them, variant size larger than 50 bp and smaller than 50,000 bp. Remained variants were visually inspected in IGV (Robinson et al., 2011) to avoid cases of false positives.

Annotation of the functional effect of the variants was done using snpEff (version 4.3t; (Cingolani et al., 2012).

BSA-seq Analysis

BSA-seq analysis was performed using the R package QTLseqr (Mansfeld and Grumet, 2018). More specifically, SNPs for the 2 bulks were first filtered using the function "filterSNPs", by keeping positions with 30 ≤ total depth ≤ 150, 0.2 ≤ reference allele frequency ≤0.8 and genotype quality ≥30. For the QTL detection, the (Takagi et al., 2013) method was applied, implemented by the function

"runQTLseqAnalysis". Briefly, SNP-index was calculated for both S- and R-bulk by dividing the number of non-reference alleles with the total number of reads in a position. If the SNP-index was <0.3 in both bulks, the SNP was discarded. SNP-index values were calculated in sliding windows of 4.7 Mbp with a 10 kb step and the average SNP-index value for each window was recorded along the chromosome. In order to avoid regions that generate segregation distortion caused by other reasons other than artificial selection (Takagi et al., 2013), the SNP-index of the S-bulk was subtracted by the SNP-index of the R-bulk in order to obtain the Δ(SNP-index). If Δ(SNP-index) equaled to 1.0 then the allele originated from the R parent and if Δ(SNP-index) equaled to -1 the allele originated from the S parent. Confidence interval calculations in each SNP position (95% and 99%), considering the null hypothesis (H_o: no QTL), was done through a simulation analysis of 10,000 replications for two bulks that were randomly generated from the population and calculating, after each iteration, the SNP-index and the corresponding Δ(SNP-index) in the two simulated bulks.

RESULTS

Phenotyping Analyses

We first analyzed symptom development and virus accumulation in P_S, P_R, and F_1 plants in a time course experiment (**Figure 1**). Symptoms started to appear in inoculated P_S plants as soon as 6 dpi (data not shown) and were conspicuous by 9 dpi, with plants showing obvious vein yellowing in all leaves and chlorotic mosaic in fully-expanded leaves (**Figure 1A**). Symptom display was delayed, and symptoms were clearly milder in P_R-infected plants, although by 9 dpi all the plants showed mild interveinal chlorotic mottling in petiole-proximal leaf areas of fully expanded leaves (**Figure 1A**). Symptoms in F_1 plants appeared at around 7 dpi, and by 9 dpi they were of intermediate severity between those of P_S- and P_R-infected plants; fully-expanded leaves showed chlorotic mottling but also mild vein yellowing, which was starting in young developing leaves (**Figure 1A**). The uniformity of symptoms in plants from each accession was remarkable, and thus each accession could be assigned to a symptom severity class without uncertainty for each observation time-point. Plants in this experiment were sampled and virus accumulation was measured, differentiating among samples from basal, medium and apical leaves (**Figure 1B**). Leaves from P_R-infected plants accumulated significantly less virus than P_S-infected leaves at all time points, except at 23 dpi in basal leaves, where no significant differences were found between both accessions. For F_1-infected plants, differences in virus accumulation in leaves were less consistent, with an apparent trend suggesting intermediate levels of accumulation between those of P_R and P_S; statistically significant differences ($P < 0.027$; Kruskal-Wallis tests) were found among the three accessions at 16 dpi in basal leaves and at 23 and 30 dpi in intermediate leaves (**Figure 1B**). Thus, virus accumulation data was essentially consistent with data on symptom expression, and phenotyping of F_1 individuals suggested incomplete dominance of the resistance trait. In any case, symptom scoring at 9 dpi appeared to be robust enough to discriminate between susceptible and resistant plants.

FIGURE 1 | Phenotyping plants of the susceptible (P_S) and resistant (P_R) parental lines, and their F_1, for CVYV susceptibility. Symptoms were scored and virus accumulation was measured at 9, 16, 23 and 30 days post-inoculation (dpi) **(A)** Symptoms in plants at 9 dpi. Symptoms could be assigned unequivocally to one of the following categories: (0) No symptoms; (1) mild chlorotic mottling in young but fully expanded leaves in interveinal petiole-proximal leaf areas; (2) similar to (1) plus vein yellowing evident in fully expanded leaves and incipient in young developing leaves; (3) obvious vein yellowing in all leaves, including young developing leaves, chlorotic mosaics in fully expanded leaves and overall plant growth reduction. **(B)** Virus accumulation was measured by quantitative dot-blot hybridization on total plant RNA extracts from basal, intermediate and apical leaves, as indicated for each graph. Three leaf discs (8 mm) were taken per leaf sampled. Virus accumulation in F_1 plants was intermediate between the susceptible and resistant parental lines in basal and intermediate leaves at 16, 23 and 30 dpi, respectively; an asterisk marks statistically significant differences in Kruskal-Wallis tests ($P < 0.027$). A minimum of 9 plants were used per treatment. Symptom category (as in (A)) is indicated for each line on the right side of each graph for each time period after inoculation.

To determine the resistance genotype of F_2 individuals, 12 individuals from each of 137 $F_{2:3}$ families were inoculated and symptom scoring performed at 9 dpi (**Supplementary Figure 1**). Out of these, 135 families were unequivocally assigned to susceptible, resistant or segregating for resistance phenotypes and, therefore, the 135 F_2 individuals could be classified as: 37 homozygous for resistance, 36 homozygous for susceptibility and 62 heterozygous, which fit well (χ^2 value 0.63, $P > 0.05$) with a 1:2:1 segregation ratio. We propose the symbol *CsCvy-1* for this monogenic incompletely-dominant resistance gene. For the BSA-seq analysis, two pools containing 15 F_2 individuals homozygous for resistance and susceptibility to CVYV, respectively, were used.

Preliminary F_2 Mapping

In parallel with the BSA-seq method, we performed F_2 mapping with a subset of the population including the *CsCvy-1* locus as a phenotypic marker. We used 172 polymorphic SNPs covering the cucumber genome: 24 SNPs in Chr01, 18 SNPs in Chr02, 34 SNPs in Chr03, 36 SNPs in Chr04, 21 SNPs in Chr05, 27 SNPs in Chr06, and 12 SNPs in Chr07 (**Supplementary Table 1**). We selected 72 F_2 individuals (35 homozygous resistant, 34 homozygous susceptible, 2 heterozygous, and 1 individual without phenotypic data) for mapping. The 72 F_2 individuals were genotyped, and most of the markers fitted the expected 1:2:1 segregation ratio. However, a group of markers on Chr05 showed a distorted segregation. This distortion was also found for the phenotypic marker *CsCvy-1*, due to the selection of individuals, with almost all of them being homozygous for this trait. The genetic map consisted of eight linkage groups (LGs) spanning 628 cM, with Chr03 split into two linkage groups (LG3A, LG3B) (**Supplementary Figure**

2, **Supplementary Table 2**). The average marker interval was 3.9 cM, with a maximum distance of 23.6 cM. The longest LG was LG4, with 115.1 cM, and the shortest was LG3A with 14 cM. The *CsCvy-1* locus was mapped onto LG05, flanked by two markers, CVYV_121 and CVYV_122 in an interval of 5.3 cM (**Figure 2A**).

In order to discard any other minor QTL involved in the resistance trait, a QTL analysis was performed. As expected, we obtained a single major QTL on LG05, co-localizing with *CsCvy-1*, with a LOD value of 49 explaining 95.8% of the variance (**Figure 2B**). No other significant QTLs were observed, in accordance to the monogenic inheritance of the trait.

Identification of *CsCvy-1* Locus by BSA-seq

Libraries of parental lines (P_S and P_R), S-bulk and R-bulk were resequenced with the Illumina HiSeq2000 sequencer. In total, 274,709,874 paired-end clean reads were mapped, after trimming and adapter removal (**Table 1**).

Small variant calling for the 4 datasets and subsequent variant filtering (see Material and Methods), generated 186,433; 215,735; 933,846 and 915,524 variants (SNPs and INDELs) for P_S, P_R, S-bulk and R-bulk, respectively, uniformly distributed throughout the genome (**Supplementary Table 3**). SNP variants from the 2 bulk datasets were used as input data in the QTLseqr R package for calculating SNP-index and Δ(SNP-index) for R- and S-bulks, based on the Takagi et al., 2013 method. Graphs for SNP-index of the R- and S-bulks and the Δ(SNP-index) were plotted (**Figure 3**). S-bulk was used as a reference dataset for the calculation of the Δ(SNP-index). BSA-seq analysis detected a single genomic region of 2,998,622 bp located in Chr05:5,088,092-8,208,448 bp, at a
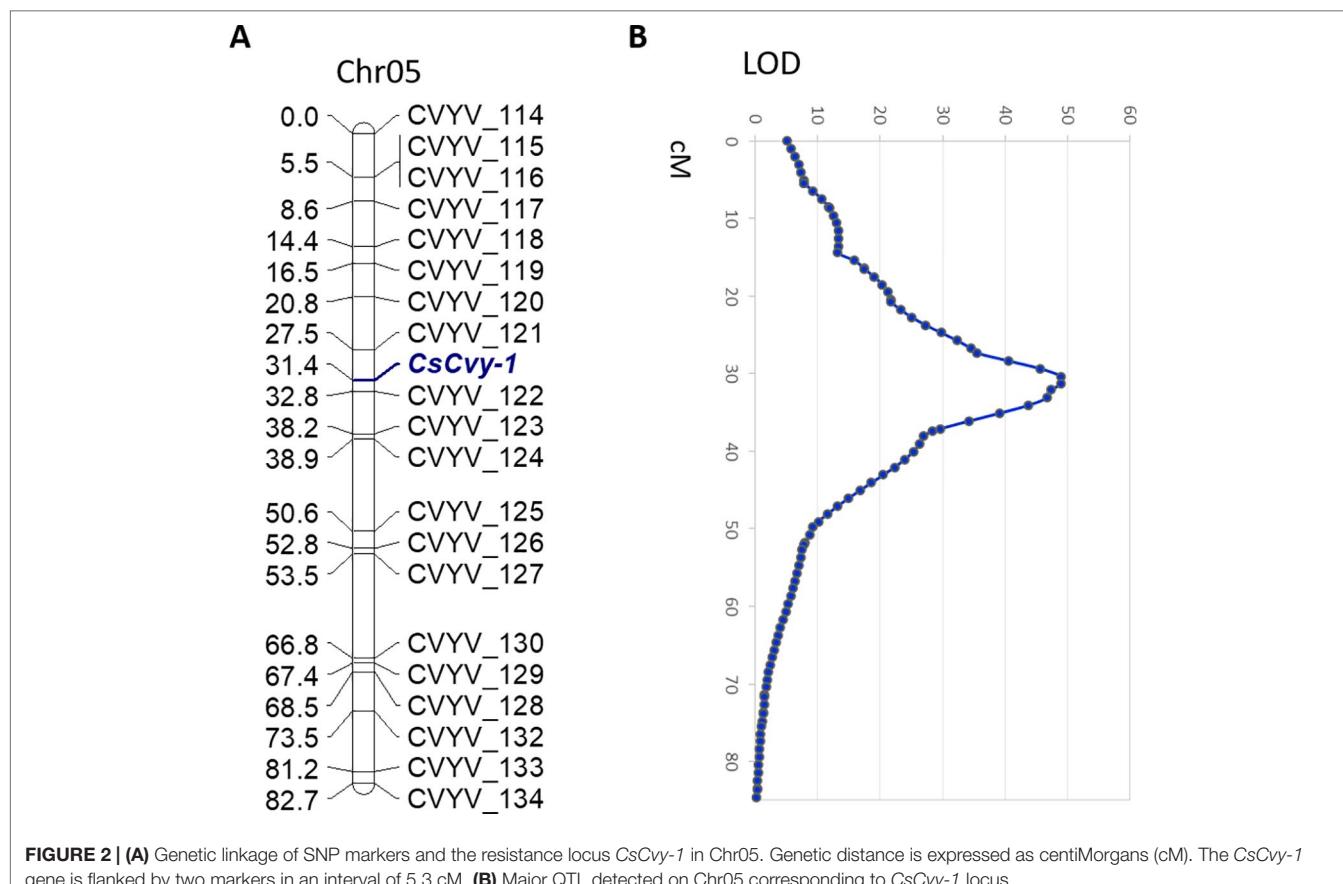
FIGURE 2 | (A) Genetic linkage of SNP markers and the resistance locus *CsCvy-1* in Chr05. Genetic distance is expressed as centiMorgans (cM). The *CsCvy-1* gene is flanked by two markers in an interval of 5.3 cM. **(B)** Major QTL detected on Chr05 corresponding to *CsCvy-1* locus.

TABLE 1 | Results from the resequencing of P_S, P_R, S-bulk and R-bulk.

Samples	Clean mapped reads[1]	Mapping rate (%)[2]	Median depth	Coverage (%)[3]
P_S	56,694,486	96	20 X	89
P_R	60,357,172	96	22 X	89
S-bulk	77,718,714	95	30 X	90
R-bulk	79,939,502	96	31 X	90

[1]Number of reads after trimming and adapter removal.
[2]Alignment to the Chinese Long 9930 genome assembly v3 (www.cucurbitgenomics.org)
[3]Coverage (≥ 1 read).

confidence interval higher than 0.99, associated with the *CsCvy-1* locus. The highest Δ(SNP-index) value was -0.528 at position 7,678,525 bp whereas the average Δ(SNP-index) was found to be -0.523 (**Figure 4A**). This region contained the flanking markers selected in the preliminary mapping.

Fine Mapping of *CsCvy-1* Locus

In order to narrow down the interval of the *CsCvy-1* locus, fine mapping was performed using the linked SNPs obtained by BSA-seq on Chr05 to design 19 SNP markers (**Figure 4B**, **Supplementary Table 1**). A subset of the F_2 population was genotyped to identify plants with recombination events, by using the genotyping data of the preliminary mapping. Eleven

recombinant informative plants delimited the candidate area of *CsCvy-1* to a 626.5 kb interval, between the markers CVYV_181 and CVYV_122 (**Figure 4C**). This interval contains 24 annotated genes (**Table 2**). To further explore the genomic region containing the *CsCvy-1* locus, we extended our previous small variant analysis with the detection of SVs between the parental lines. In total, we detected 2 SVs of more than 50 bp in length: one 55 bp deletion and a duplication measuring 41,644 bp (**Supplementary Table 4**).

With the purpose of detecting alterations in potential candidate genes that could be responsible for the resistance phenotype, we performed an annotation of the effect of small variations in the 24 genes found in the *CsCvy-1* region. In total, we annotated 45 SNPs/INDELs, of which 17 were highlighted for being within or close to coding sequences of 10 genes (**Supplementary Table 5**). The vast majority of variations (14 out of 17) were non-synonymous, and were annotated as modifiers or with low impact. Among the remaining four variants, three SNPs caused synonymous changes, and one insertion of 2 nucleotides caused a frameshift resulting in a premature stop codon (**Supplementary Table 5**). The three missense variants (moderate impact) were located in the coding sequence of CsaV3_5G011200, CsaV3_5G011220 and CsaV3_5G011240 genes, whereas the frameshift variant (high impact) was located in CsaV3_5G11180 gene. The frameshift caused the disruption of the CsaV3_5G11180 gene, coding for a serine/arginine repetitive matrix protein (SARMP) 2-like, at position 607 (**Supplementary Figure 3**). A search for
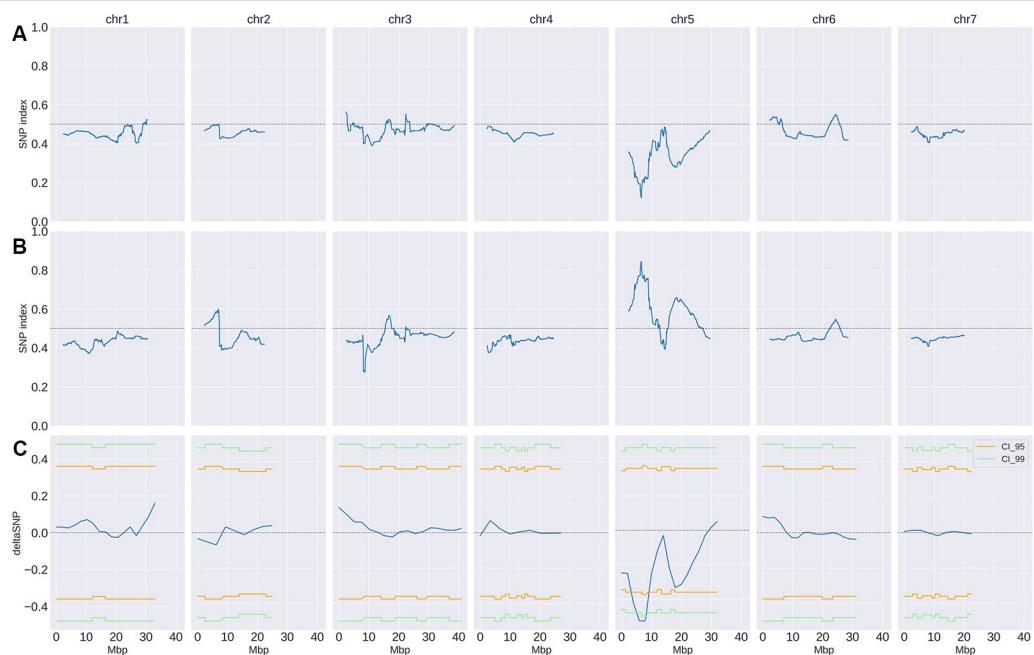
FIGURE 3 | SNP-index and Δ(SNP-index) distribution (blue line). SNP-index for R-bulk **(A)** and S-bulk **(B)** was calculated using sliding windows of 4.7 Mbp in length with a step measuring 10 kb. The corresponding Δ(SNP-index) **(C)** was calculated as the difference SNP-index R-bulk − SNP-index S-bulk. Regions of statistical significance are detected as those that surpass the threshold of 0.95 (orange line) or 0.99 (green line) confidence intervals.

TABLE 2 | List of genes within the interval of *CsCVY-1* locus in the Chinese Long 9930 v3 cucumber genome annotation, variants and structural variation analysis between P_S and P_R.

Gene_ID	Description Chinese Long v3	Small variants[1]
CsaV3_5G011170*	Unknown protein	−
CsaV3_5G011180*	Serine/arginine repetitive matrix protein 2-like isoform X1	INDEL
CsaV3_5G011190*	Unknown protein	−
CsaV3_5G011200*	RNA-dependent RNA polymerase	SNP
CsaV3_5G011210*	RNA-dependent RNA polymerase	SNP
CsaV3_5G011220*	Endo-1,4-beta-xylanase A-like	SNP
CsaV3_5G011230	Unknown protein	SNP
CsaV3_5G011240	SAM dependent carboxyl methyltransferase	SNP
CsaV3_5G011250	Unknown protein	SNP
CsaV3_5G011260	SAM dependent carboxyl methyltransferase	SNP
CsaV3_5G011270	Unknown protein	−
CsaV3_5G011280	Integrator complex subunit 9 like	SNP
CsaV3_5G011290	Cytochrome c	−
CsaV3_5G011300	Serine/threonine protein kinase	SNP
CsaV3_5G012300	40S ribosomal protein S6	−
CsaV3_5G012310	Myb family transcription factor	−
CsaV3_5G012320	Protein kinase, putative	−
CsaV3_5G012330	E3 ubiquitin-protein ligase ORTHRUS 2-like	−
CsaV3_5G012340	THUMP domain-containing protein 1	−
CsaV3_5G012350	Amino acid transporter	−
CsaV3_5G012360	Unknown protein	−
CsaV3_5G012370	Perakine reductase	−
CsaV3_5G012380	UBN2_3 domain-containing protein	−
CsaV3_5G012390	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1	−

[1]Small variants are referred as SNPs or INDELs < 50 bp.
*These genes are inside the duplicated region of 41,644 bp in P_R.

conserved motifs detected a Constitutive Photomorphogenic 1 (COP1)-interacting protein signature. With regards to structural variations, we also detected a single duplication event in P_R measuring 41,644 bp in length located in Chr05:7,195,565-7,237,209 that contained genes CsaV3_5G011170 to CsaV3_5G011220 (**Table 2** and **Supplementary Figure 4**). Genes CsaV3_5G011170 and CsaV3_5G011190 encode for unknown proteins, CsaV3_5G11180 encodes the above-mentioned SARMP and genes CsaV3_5G011200 and CsaV3_5G011210 encode two RNA-dependent RNA polymerases (RDRs) 1a and 1b, respectively. In contrast, the deletion of 55 bp was located in an intergenic region.

DISCUSSION

In cucumber, the disease caused by CVYV is a limiting factor for production in areas with high pressure from viruliferous whiteflies. In this work, we phenotyped an $F_{2:3}$ segregating population identifying a single monogenic locus, *CsCvy-1*, that controls resistance to CVYV. The resistance conferred by this locus is partial, as P_R plants showed viral accumulation in systemically infected leaves. However, the progression of the disease was very much reduced as compared to susceptible controls, the growth of the plants was not affected at all, and viral accumulation was significantly reduced as compared to P_S and F_1 plants. In cucumber, the accession C.sat-10 (Picó et al., 2003) was described as having partial resistance against CVYV, and a segregating F_2 population was obtained to study the genetic control of this resistance (Picó et al., 2008); the segregation fitted
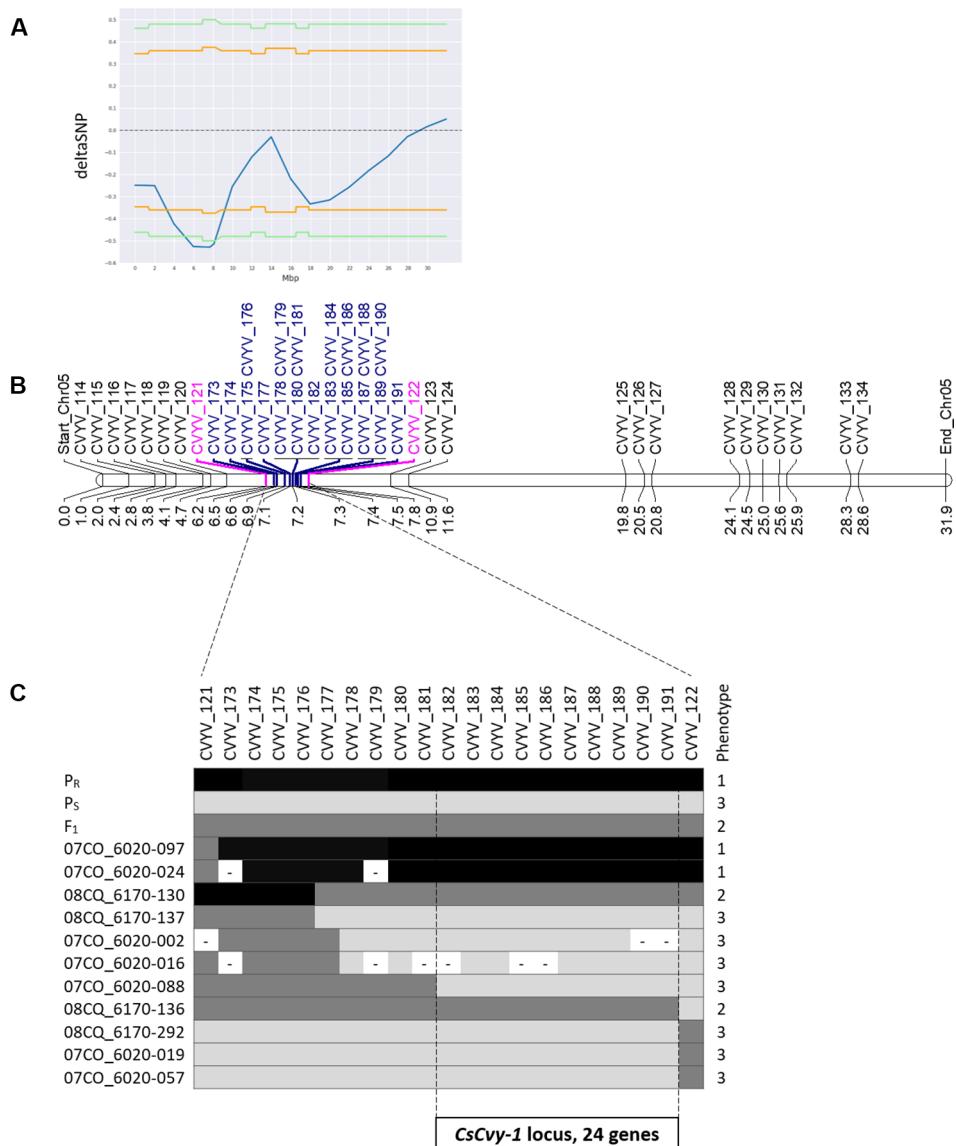
FIGURE 4 | Scheme of fine mapping of the *CsCvy-1* locus. **(A)** DeltaSNP graph on Chr05 with the statistical confidence interval of 0.95 (orange line) or 0.99 (green line) **(B)** Physical map of Chr05 with all the markers used for mapping the *CsCvy-1* locus: preliminary mapping (black), flanking markers (magenta), fine mapping markers (blue). Distance is expressed in Mb on the bottom of the bar. **(C)** Genotyping result of recombinant individuals with the markers covering *CsCvy-1* locus. Alleles are represented in three colors: black for homozygous P_R allele, dark grey for heterozygous P_R/P_S, and light grey for homozygous P_S allele.

a monogenic control with dominance. These features are very similar to what we have described here; however, in the case of C.sat-10, no mapping studies were performed to determine the localization of the locus in the cucumber genome. Thus, although our resistance data fit well with those from Picó et al. (2003; 2008), we cannot rule out that both resistances to CVYV characterized in cucumber are independent. By comparing with closely related species such as melon, Pitrat et al. (2012) evaluated a collection of 1,188 accessions for resistance against CVYV, and studied the inheritance in F_1, F_2 and BC progenies. Three loci were detected in their work: *Cvy-1*, controlling resistance in PI 164323 and necrosis in HSD 93-20-A; *cvy-2*, showing recessive

tolerance in HSD 2458, and *Cvy-3*, showing dominance for severe mosaic symptoms in Ouzbèque 2.

Breeding for resistance requires the availability of highly linked markers that can be utilized for performing rapid and specific introgressions of the desired trait. For marker development, conventional gene mapping is based on the phenotyping and genotyping of a large number of individuals in a population, and it is time-consuming, costly and limiting in terms of the size of the population (Takagi et al., 2013). One way to improve conventional mapping is to use BSA-seq, in which the number of individuals to be analyzed can be narrowed down to two representative bulks, and at the same time as the mapping, the resequencing data offers

the possibility of obtaining a high number of markers linked to the trait (Yang et al., 2013). BSA-seq has been successfully applied for the mapping of important agronomical traits in many crops such as rice (Abe et al., 2012; Takagi et al., 2013; Yang et al., 2013; Sun et al., 2018), lettuce (Huo et al., 2016), potato (Kaminski et al., 2016), soybean (Song et al., 2017), broccoli (Shu et al., 2018; Branham and Farnham, 2019) or sorghum (Han et al., 2015). Moreover, in cucurbits BSA-seq has enabled the identification of candidate genes for dwarfism (Dong et al., 2018), yellow skin (Dou et al., 2018) and light rind color (Oren et al., 2019) in watermelon; mapping flavor traits in melon (Zhang et al., 2016); or the identification of candidate genes for flesh thickness (Xu et al., 2015), aphid resistance (Liang et al., 2016), early flowering QTL (Lu et al., 2014), two major QTLs for downy mildew resistance (Win et al., 2017), and three major QTLs conferring subgynoecy (Win et al., 2019) in cucumber. In the present study, through the use of a BSA-seq strategy, the *CsCvy-1* locus has been successfully mapped to a region of 2.9 Mb in chromosome 5, whereas no other regions of the genome exhibited significant association with the resistance. These results confirmed previous analysis performed with the conventional mapping approach and were later used for fine mapping. The BSA-seq analysis provided enough SNPs to fine map the trait within a narrow interval of 625 kb, containing only 24 annotated genes.

In order to identify candidate genes for *CsCvy-1*, we performed an analysis of small variants and structural variation around this locus. Most of the variations had a small, if any, predicted impact, except for the insertion of 2 nucleotides in the gene encoding a SARMP2-like, which causes a frameshift mutation in the coding sequence with the subsequent truncation of the protein. A functional analysis of the SARMP-like protein sequence detected a COP1-interacting protein signature. Interestingly, COP1 has been associated with plant pathogen resistance (Lim et al., 2018) and the *Arabidopsis thaliana* COP1-interacting protein is a positive regulator of ABA response (Ren et al., 2016), which is an essential regulator of plant immunity (Berens et al., 2017). Nevertheless, perhaps the most appealing modification within the *CsCvy-1* region is the duplication of the 41 Kb fragment containing the genes encoding RDRs 1a and 1b (CsaV3_5G011200 and CsaV3_5G011210). RDRs are critical players in RNA silencing pathways; they are the key enzymes in the process of amplification of double-stranded RNAs that activate gene silencing after nuclease processing. A role for RDRs in antiviral immunity has long been acknowledged, in particular for members of the RDR1 and RDR6 clades (Willmann et al., 2012). For instance, the absence of a functional RDR1 in *Nicotiana benthamiana* can explain enhanced susceptibility to many viruses in this species (Yang et al., 2004). In relation to virus resistance in crop species, it has been recently demonstrated that the tomato genes *Ty-1* and *Ty-3* for resistance to tomato yellow leaf curl virus are alleles of an RDR gene (Verlaan et al., 2013). In cucumber, a recent report shows that the RDR1a and 1b genes have enhanced expression in natural or engineered lines showing broad virus resistance; importantly, one of the viruses tested was CVYV (Leibman et al., 2018). Taking these data together, a mechanistic explanation for our observations may consist of

enhanced antiviral activity in the P_R line as a consequence of enhanced RDR1a and/or 1b expression; this, in turn, would be the consequence of the described gene duplication. This is an attractive hypothesis that awaits further testing, although at least a good alternative candidate gene (*i.e.,* the gene encoding a SARMP2-like protein) exists. From the point of view of resistance stability when confronted to different CVYV strains (Desbiez et al., 2019), the potential implication of RDR1a/b represents an optimistic perspective, given its broad spectrum of action (Leibman et al., 2018).

In conclusion, in this work we identified the monogenic locus *CsCvy-1*, inherited under incomplete dominance, in a short interval of 5.3 cM containing 24 genes. This is the first report where a CVYV resistance locus has been mapped in cucumber, and valuable molecular markers for MAS breeding programs have been developed. Moreover, our findings will be the basis for further map-based cloning and functional validation of the resistance gene.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher. Variation data were deposited to European Variation Archive (EVA: https://www.ebi.ac.uk/eva) with the project accession number PRJEB34274.

AUTHOR CONTRIBUTIONS

MA, JG-M, TJ, and MP conceived and designed the research. PM, MM and MA provided the plant material and performed the tests with CVYV. A-SF conducted the conventional mapping. KA performed the bioinformatics analysis of the BSA-seq. MP conducted marker development, mapping analysis, and wrote the manuscript with important contributions from MA and KA. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01583/full#supplementary-material

REFERENCES

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

Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T., and Tsuda, K. (2017). Evolution of hormone signaling networks in plant defense. *Annu. Rev. Phytopathol.* 55, 401–425. doi: 10.1146/annurev-phyto-080516-035544

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

Branham, S. E., and Farnham, M. W. (2019). Identification of heat tolerance loci in broccoli through bulked segregant analysis using whole genome resequencing. *Euphytica* 215, 1–9. doi: 10.1007/s10681-018-2334-9

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

Crow, J. F. (1990). Mapping functions. *Genetics* 125, 669–671.

Desbiez, C., Caciagli, P., Wipf-Scheibel, C., Millot, P., Ruiz, L., Marian, et al. (2019). Evidence for long-term prevalence of cucumber vein yellowing virus in Sudan and genetic variation of the virus in Sudan and the Mediterranean Basin. *Plant Pathol.* 68, 1268–1275. doi: 10.1111/ppa.13055

Dong, W., Wu, D., Li, G., Wu, D., and Wang, Z. (2018). Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Sci. Rep.* 8, 1–7. doi: 10.1038/s41598-018-21293-1

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

Doyle, J. (1991). "DNA Protocols for Plants," in *Molecular Techniques in Taxonomy. NATO ASI Series (Series H: Cell Biology)*, vol. 57. Eds. G. M. Hewitt, A. W. B. Johnston, and J. P. W. Young (Berlin, Heidelberg: Springer). doi: 10.1007/978-3-642-83962-7_18

Galipienso, L., Janssen, D., Rubio, L., Aramburu, J., and Velasco, L. (2013). Cucumber vein yellowing virus isolate-specific expression of symptoms and viral RNA accumulation in susceptible and resistant cucumber cultivars. *Crop Prot.* 43, 141–145. doi: 10.1016/j.cropro.2012.08.004

Han, Y., Lv, P., Hou, S., Li, S., Ji, G., Ma, X., et al. (2015). Combining next generation sequencing with bulked segregant analysis to fine map a stem moisture locus in sorghum (Sorghum bicolor L. *Moench*). *PloS One* 10, 1–14. doi: 10.1371/journal.pone.0127065

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, 1275–1281. doi: 10.1038/ng.475

Huo, H., Henry, I. M., Coppoolse, E. R., Verhoef-Post, M., Schut, J. W., de Rooij, H., et al. (2016). Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing. *Plant J.* 88, 345–360. doi: 10.1111/tpj.13267

Janssen, D., Velasco, L., Martín, G., Segundo, E., and Cuadrado, I. M. (2007). Low genetic diversity among Cucumber vein yellowing virus isolates from Spain. *Virus Genes* 34, 367–371. doi: 10.1007/s11262-006-0026-3

Kaminski, K. P., Kørup, K., Andersen, M. N., Sønderkær, M., Andersen, M. S., Kirk, H. G., et al. (2016). Next generation sequencing bulk segregant analysis of potato support that differential flux into the cholesterol and stigmasterol metabolite pools is important for steroidal glycoalkaloid content. *Potato Res.* 59, 81–97. doi: 10.1007/s11540-015-9314-4

Leibman, D., Kravchik, M., Wolf, D., Haviv, S., Weissberg, M., Ophir, R., et al. (2018). Differential expression of cucumber RNA-dependent RNA polymerase 1 genes during antiviral defence and resistance. *Mol. Plant Pathol.* 19, 300–312. doi: 10.1111/mpp.12518

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Liang, D., Chen, M., Qi, X., Xu, Q., Zhou, F., and Chen, X. (2016). QTL Mapping by SLAF-seq and expression analysis of candidate genes for aphid resistance in cucumber. *Front. Plant Sci.* 7, 1–8. doi: 10.3389/fpls.2016.01000

Lim, G.-H., Hoey, T., Zhu, S., Clavel, M., Yu, K., Navarre, D., et al. (2018). COP1, a negative regulator of photomorphogenesis, positively regulates plant disease resistance via double-stranded RNA binding proteins. *PloS Pathog.* 14, e1006894. doi: 10.1371/journal.ppat.1006894

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. *Theor. Appl. Genet.* 127, 1491–1499. doi: 10.1007/s00122-014-2313-z

Mansfeld, B. N., and Grumet, R. (2018). QTLseqr: an R package for bulk segregant analysis with next-generation sequencing. *Plant Genome* 11, 0. doi: 10.3835/plantgenome2018.01.0006

Marco, C. F., Aguilar, J. M., Abad, J., Gómez-Guillamón, M. L., and Aranda, M. A. (2003). Melon Resistance to Cucurbit yellow stunting disorder virus Is Characterized by Reduced Virus Accumulation. *Phytopathology* 93, 844–852. doi: 10.1094/phyto.2003.93.7.844

Martínez-García, B., Marco, C. F., Goytia, E., López-Abella, D., Serra, M. T., Aranda, M. A., et al. (2004). Development and use of detection methods specific for Cucumber vein yellowing virus (CVYV). *Eur. J. Plant Pathol.* 110, 811–821. doi: 10.1007/s10658-004-2491-7

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

Navas-Castillo, J., López-Moya, J. J., and Aranda, M. A. (2014). Whitefly-transmitted RNA viruses that affect intensive vegetable production. *Ann. Appl. Biol.* 165, 155–171. doi: 10.1111/aab.12147

Oren, E., Tzuri, G., Vexler, L., Dafna, A., Meir, A., Faigenboim, A., et al. (2019). Multi-allelic APRR2 Gene is associated with fruit pigment accumulation in melon and watermelon. *J. Exp. Bot.* 70, 3781–3798. doi: 10.1093/jxb/erz182

Picó, B., Villar, C., and Nuez, F. (2003). Screening Cucumis sativus landraces for resistance to cucumber vein yellowing virus. *Plant Breed.* 122, 426–430. doi: 10.1046/j.1439-0523.2003.00882.x

Picó, B., Sifres, A., Martinez-Perez, E., Leiva-Brondo, M., and Nuez, F. (2008). Genetics of the resistance to CVYV in cucumber. In Modern variety breeding for present and future needs. *Proceedings of the 18th EUCARPIA general congress*, (J Prohens, and ML Badenes, eds), Valencia (Spain), pp 452–456.

Pitrat, M., Wipf-Scheibel, C., Besombes, D., Desbiez, C., and Lecoq, H. (2012). Resistance of melon to Cucumber Vein Yellowing Virus (CVYV). *Cucurbitaceae 2012, Proceedings of the Xth EUCARPIA Meeting on Genetics and Breeding of Cucurbitaceae* (eds. Sari, Solmaz, and Aras) Antalya (Turkey) pp 157–164.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., and Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, 333–339. doi: 10.1093/bioinformatics/bts378

Ren, Y., Zhang, Z., Liu, J., Staub, J. E., Han, Y., Cheng, Z., et al. (2009). An integrated genetic and cytogenetic map of the cucumber genome. *PLoS ONE* 4 (6), e5795. doi: 10.1371/journal.pone.0005795

Ren, C., Zhu, X., Zhang, P., and Gong, Q. (2016). Arabidopsis COP1-interacting protein 1 is a positive regulator of ABA response. *Biochem. Biophys. Res. Commun.* 477, 847–853. doi: 10.1016/j.bbrc.2016.06.147

Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E., Getz, G., et al. (2011). Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Shu, J., Liu, Y., Zhang, L., Li, Z., Fang, Z., Yang, L., et al. (2018). QTL-seq for rapid identification of candidate genes for flowering time in broccoli × cabbage. *Theor. Appl. Genet.* 131, 917–928. doi: 10.1007/s00122-017-3047-5

Song, J., Li, Z., Liu, Z., Guo, Y., and Qiu, L.-J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Front. Plant Sci.* 8, 1–11. doi: 10.3389/fpls.2017.00919

Staub, J. E., Robbins, M. D., and Wehner, T. C. (2008). "Cucumber," in *Vegetables I. Handbook of Plant Breeding*, vol. 1 . Eds. J. Prohens, and F. Nuez (New York, NY: Springer). doi: 10.1007/978-0-387-30443-4_8

Sun, Y., Wang, J., Crouch, J. H., and Xu, Y. (2010). Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. *Mol. Breed.* 26, 493–511. doi: 10.1007/s11032-010-9390-8

Sun, J., Yang, L., Wang, J., Liu, H., Zheng, H., Xie, D., et al. (2018). Identification of a cold-tolerant locus in rice (Oryza sativa L.) using bulked segregant analysis with a next-generation sequencing strategy. *Rice* 11, 1–12. doi: 10.1186/s12284-018-0218-1

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome

resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Verlaan, M. G., Hutton, S. F., Ibrahem, R. M., Kormelink, R., Visser, R. G. F., Scott, J. W., et al. (2013). The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 Are Allelic and Code for DFDGD-Class RNA-Dependent RNA Polymerases. *PLoS Genet.* 9 (3), e1003399. doi: 10.1371/journal.pgen.1003399

Wóycicki, R., Witkowicz, J., Gawroński, P., Dabrowska, J., Lomsadze, A., Pawełkowicz, M., et al. (2011). The genome sequence of the North-European Cucumber (Cucumis sativus l.) unravels evolutionary adaptation mechanisms in plants. *PLoS ONE* 6 (7), e22728. doi: 10.1371/journal.pone.0022728

Willmann, M. R., Endres, M. W., Cook, R. T., and Gregory, B. D. (2012). The functions of RNA-dependent RNA polymerases in Arabidopsis. *Arab. B.* 9, e0146. doi: 10.1199/tab.0146

Win, K. T., Vegas, J., Zhang, C., Song, K., and Lee, S. (2017). QTL mapping for downy mildew resistance in cucumber via bulked segregant analysis using next-generation sequencing and conventional methods. *Theor. Appl. Genet.* 130, 199–211. doi: 10.1007/s00122-016-2806-z

Win, K. T., Zhang, C., Silva, R. R., Lee, J. H., Kim, Y. C., and Lee, S. (2019). Identification of quantitative trait loci governing subgynoecy in cucumber. *Theor. Appl. Genet.* 132, 1505–1521. doi: 10.1007/s00122-019-03295-3

Xu, X., Lu, L., Zhu, B., Xu, Q., Qi, X., and Chen, X. (2015). QTL mapping of cucumber fruit flesh thickness by SLAF-seq. *Sci. Rep.* 5, 1–9. doi: 10.1038/srep15829

Yang, S.-J., Carter, S. A., Cole, A. B., Cheng, N.-H., and Nelson, R. S. (2004). A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by Nicotiana benthamiana. *Proc. Natl. Acad. Sci.* 101, 6297–6302. doi: 10.1073/pnas.0304346101

Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A. J., et al. (2013). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS ONE* 8 (7), e68433. doi: 10.1371/journal.pone.0068433

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871. doi: 10.1093/bioinformatics/btp394

Zhang, H., Yi, H., Wu, M., Zhang, Y., Zhang, X., Li, M., et al. (2016). Mapping the flavor contributing traits on "Fengwei melon" (Cucumis melo L.) chromosomes using parent resequencing and super bulked-segregant analysis. *PloS One* 11, 1–24. doi: 10.1371/journal.pone.0148150

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

Quantitative Trait Loci Mapping and Candidate Gene Analysis of Low Temperature Tolerance in Cucumber Seedlings

Shaoyun Dong†, Weiping Wang†, Kailiang Bo†, Han Miao, Zichao Song, Shuang Wei, Shengping Zhang* and Xingfang Gu*

Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

Cucumber (*Cucumis sativus* L.) is an economically important vegetable crop worldwide, but it is sensitive to low temperatures. Cucumber seedlings exposed to long-term low temperature stress (LT), i.e., below 20°C during the day, and 8°C at night, exhibit leaf yellowing, accelerated senescence, and reduced yield, therefore posing a threat to cucumber production. Studying the underlying mechanisms involved in LT tolerance in cucumber seedlings, and developing germplasm with improved LT-tolerance could provide fundamental solutions to the problem. In this study, an F_2 population was generated from two parental lines, "CG104" (LT-tolerant inbred line) and "CG37" (LT-sensitive inbred line), to identify loci that are responsible for LT tolerance in cucumber seedlings. Replicated phenotypic analysis of the F_2-derived F_3 family using a low-temperature injury index (LTII) suggested that the LT tolerance of cucumber seedlings is controlled by multiple genes. A genetic map of 990.8 cM was constructed, with an average interval between markers of 5.2 cM. One quantitative trait loci (QTL) named *qLTT5.1* on chromosome 5, and two QTLs named *qLTT6.1* and *qLTT6.2* on chromosome 6 were detected. Among them, *qLTT6.2* accounted for 26.8 and 24.1% of the phenotypic variation in two different experiments. Single-nucleotide polymorphism (SNP) variations within the region of *qLTT6.2* were analyzed using two contrasting *in silico* bulks generated from the cucumber core germplasm. Result showed that 214 SNPs were distributed within the 42-kb interval, containing three candidate genes. Real-time quantitative reverse transcription PCR and sequence analysis suggested that two genes *Csa6G445210*, an auxin response factor, and *Csa6G445230*, an ethylene-responsive transmembrane protein, might be candidate genes responsible for LT tolerance in cucumber seedlings. This study furthers the understanding of the molecular mechanism underlying LT tolerance in cucumber seedlings, and provides new markers for molecular breeding.

Keywords: cucumber, low temperature tolerance, quantitative trait locus mapping, *in silico* bulked segregant analysis, candidate gene analysis

INTRODUCTION

Cucumber (*Cucumis sativus* L.) is an economically important vegetable crop. Global production reached 80 million metric tons in 2016, with a steady increase in production since 1998 (Chen et al., 2019). Cucumber originated from tropical regions, and the suitable temperature range for growth is 18–30°C. As an annual vegetable crop which is sensitive to low temperature (LT) (Cabrera et al., 1992), cucumber easily suffers from LT injury in winter and early spring in regions such as Northern China, necessitating greenhouse-cultivation in these regions. Depending on its intensity and duration, LT-stress may be divided into long-term moderate LT (below 20°C during daytime and 8°C at night for days or weeks) and short-term extreme LT (15°C during the day and 4°C at night for hours) (Smeets and Wehner, 1997). Long-term moderate LT stress occurs more often during production, leading to yellowed leaves, accelerated senescence, and decreased yield (Smeets and Wehner, 1997), such that even this moderate stress has become a serious obstacle for the cucumber production industry. Thus, LT tolerance at the seedling stage is a desirable trait for cucumber breeding. With the continuous expansion of the cucumber planting area into early spring and winter in China, it is critical to identify candidate genes responsible for LT tolerance, and to develop resistant cultivars.

LT tolerance of cucumber seedling is a complex trait controlled by multiple alleles at multiple loci (Chung et al., 2003; Kozik and Wehner, 2008; Kozik et al., 2012). Chung et al. (2003) found that LT tolerance at the seedling stage is maternally inherited, and that the gene conferring resistance comes from the chloroplast genome of the female parent "Chipper." Kozik and Wehner (2008) reported that the LT tolerance of cucumber seedlings was controlled by a single dominant gene *Ch*. More recently, a study of two genotypes, LT-tolerant "PI390953" and cold-sensitive "Gy14" suggested that LT tolerance of cucumber seedlings was determined by two genes (Kozik et al., 2012).

There are few recent reports on the gene or quantitative trait loci (QTL) mapping of LT tolerance in cucumber seedlings. A few studies were carried out under short-term extreme LT (Li, 2014; Wang, 2014; Zhou, 2015). When a qualitative chilling injury index (CII) was used as an indicator of LT stress, one simple sequence repeat (SSR) marker closely linked to a LT-tolerant gene on chromosome 6 was identified (Li, 2014). Zhou (2015) identified six loci associated with LT tolerance on chromosome 3. Wang (2014) used CII and a recovery index as indicators of LT tolerance, and identified three QTLs on chromosome 3 and one QTL on chromosome 7, respectively. However, there are no reports of LT-tolerant QTLs or gene mapping using long-term moderate LT as a selective criterion, and long-term moderate LT stress occurs more often during cucumber production.

Since the International Cucumber Genome Team announced the whole genome sequence of "Chinese long" line 9930 (Huang et al., 2009), multiple lines have been re-sequenced, revealing all genotypic variability (Qi et al., 2013). A new approach called "*in silico* BSA (bulked segregant analysis)" was reported (Bello et al., 2014). Its application avoids the construction of near isogenic lines and greatly improves the efficiency of gene mapping. The method is based on the results of gene/QTL mapping and

genetically diverse germplasm with known phenotype. Two bulks with contrasting target phenotype are constructed, and the single-nucleotide polymorphism (SNP) variations within the target QTL region between two bulks are inspected. The SNPs that are identical within each bulk but different between two bulks may be highly associated with the target phenotype. With this method, Li et al. (2016) quickly narrowed the *Cn* locus into a 16-kb region, and Bo et al. (2019) fine-mapped the fruit spine density related major QTL *qfsd6.2* to a 50-kb region, and finally identified the candidate gene *Csgl3*.

So far, studies on LT tolerance in cucumber mainly focused on the LT stress conditions applied, genetic mechanisms and germplasm evaluation criteria (Kozik and Wehner, 2008; Kozik et al., 2012). However, very few QTL analyses of LT tolerance in cucumber have been reported and no genes have been identified. Such information would provide practical tools to enhance breeding strategies for this trait. Therefore, the objective of this study was to identify QTL and candidate genes responsible for LT stress tolerance. Our preliminary study showed that the cucumber inbred line "CG104" exhibits high tolerance to LT, while inbred line "CG37" is sensitive to LT. An F_2 population from a cross of these two inbred lines was constructed, and used for QTL mapping. Then, SNP variations within the major QTL region between two contrasting *in silico* bulks were analyzed. Sequence analysis and quantitative reverse transcription PCR analysis were then performed to identify the potential candidate gene. This study thus helps to further the understanding of the molecular mechanism underlying LT tolerance in cucumber seedlings, and provides new tools for breeding LT-tolerant cucumber germplasm *via* molecular marker-assisted breeding.

MATERIALS AND METHODS

Plant Materials

Two cucumber inbred lines, "CG104" (LT-tolerant) and "CG37" (LT-sensitive) were used in this study. "CG37" and "CG104" were crossed to generate an F_1 ("CG104" x "CG37") and F_1' ("CG37" x "CG104") populations. The F_1 ("CG104" x "CG37") was self-pollinated to produce an F_2 population of 189 individuals. A total of 189 $F_{2:3}$ families were derived from F_2 populations accordingly. Moreover, 10 core germplasms with high LT-sensitive or LT-resistant phenotype (**Table S1**) that have been resequenced (Qi et al., 2013) were used for *in silico* BSA analysis. The F_2 population was used to verify mutation sites in candidate genes. All materials were preserved by the Cucumber Research Group, Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS).

Investigation of Low Temperature Injury at Seedling Stage

LT treatments were carried out twice in the greenhouse and plastic greenhouse of the IVF, CAAS at Changping, Beijing (40°13'N, 116°05'E) in early spring, respectively (**Table S2**). Three-week-old seedlings of the two parental lines, F_1 and F_2-derived F_3 families were exposed to 15–17°C for 2 weeks, and the LT injury

was classified into six grades, based on the degree of yellowing and dryness of the cotyledons and the first true leaf (**Figure 1**). The index used was as follows: 0: no symptoms on either the cotyledons or the first true leaf; 1: cotyledons were slightly yellow, while the first true leaf was still green; 3: cotyledons were yellow-green with yellowed edges, while the first true leaf was light green; 5: cotyledons were yellowed in large scale, while the first true leaf was yellow-green; 7: cotyledons were completely

FIGURE 1 | Phenotypic characterization of low temperature tolerance in $F_{2:3}$ populations. **(A)** Scales of low-temperature injury index (LTII) based on the degree of yellowing and dryness of the cotyledons and the first true leaf. 0: no symptoms on either the cotyledons or first true leaf; 1: cotyledons were slightly yellow, while the first true leaf was still green; 3: cotyledons were yellow-green with yellowed edges, while the first true leaf was light green; 5: cotyledons were yellowed in large scale, while the first true leaf was yellow-green; 7: cotyledons were completely yellowed, while the first true leaf was yellow-green; 9: cotyledons and the first true leaf dried. **(B)** Frequency distribution of LTII in F_2 populations characterized in March 2017 and April 2017.

yellowed, while the first true leaf was yellow-green; 9: cotyledons and the first true leaf dried.

A low-temperature injury index (LTII) was used as an indicator to indicate the LT tolerance of each plant. The formula used for the calculation of LTII refers to Wang et al. (2019): LTII = $(0×S_0+1×S_1+3×S_3+5×S_5+7×S_7+9×S_9)/N×9$. S_0–S_9 indicates the number of plants corresponding to each grade. N indicates the total number of plants. Three replicates were set for each treatment, and eight plants of each replicate were investigated. The LTII in two experiments was evaluated by two people, respectively.

Deoxyribonucleic Acid Extraction and Simple Sequence Repeats Marker Analysis

A modified CTAB (cetyltrimethylammonium ammonium bromide) procedure (Wang et al., 2006) was applied to extract genomic DNA from the second leaf of each two-leaf stage seedlings. The DNA concentration and quality was examined by electrophoresis on a 1% (w/v) agarose gel and NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). A total of 1,288 SSR markers designed based on the genome sequence of cucumber (Ren et al., 2009) were used to screen the polymorphism of the two parents. Then, a linkage map was constructed with selected polymorphic markers. PCR amplification and gel electrophoresis were conducted referring to the method of Song et al. (2016).

Linkage Map Construction and Quantitative Trait Locus Mapping

Polymorphic primers were used for genotyping F_2 individuals. QTL analysis was performed with the R/QTL software package (http://www.rqtl.org/). Using individual plant data, a whole genome scan was performed to map the QTLs with composite interval mapping (CIM) procedures (Weng et al., 2015). Genome-wide logarithm of the odds (LOD), threshold values ($P < 0.05$) for declaring the presence of QTLs were determined using 1,000 permutations. For each detected QTL, a 2-LOD support interval was calculated and defined by left and right markers. The QTL naming format is the abbreviation of the character (i.e., low temperature tolerance—LTT), chromosome (Chr.) number and locus number.

Defining the Interval Containing the Major Quantitative Trait Locus Using *In Silico* Bulked Segregant Analysis

To narrow down the region harboring the major QTL, the association between SNPs and LT tolerance in core germplasm was studied using the *in silico* BSA strategy. The LT tolerance of 87 sequenced core germplasm (CG) seedlings was previously evaluated in our lab (Wang et al., 2019), two-leaf stage seedlings of CG lines were exposed to 12°C for 11 d and 19.3°C for 14 d, respectively. The LTT of each line was evaluated using the same method as that used to phenotype the $F_{2:3}$ families described above. Five highly LT-resistant CG lines ("CG29," "CG56,"

"CG61," "CG90," "CG104") and five highly LT-sensitive CG lines ("CG10," "CG21," "CG43," "CG109," "CG37") were selected to generate LT-resistant and LT-sensitive bulks (**Table S1**, **Figure S2**). The 10 CG lines have different geographical origin, and were re-sequenced (Qi et al., 2013). The sequence data was retrieved from National Center for Biotechnology Information (NCBI) Short Read Archive (SRA; Leionen et al., 2011) under accession SRA056480 (https://www.ncbi.nlm.nih.gov/sra/?term=SRA056480). The detailed accession number of each CG line was listed in **Table S1**. Then, the SNPs located within the *qLTT6.2.* region (20,591,185 to 21,186,690 bp) were searched (**Table S3**) and imported into PilotEdit software. SNPs that were identical *within* each bulk while different *between* bulks, were considered to be associated with the LT tolerance.

Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction Analysis

The parental lines at the two-leaf stage were exposed to 14 h light at 18°C and 10 h dark at 10°C, the second true leaf of each seedling was harvested at 0, 10, 24, 48, 96, 144, 216, 288, 384, and 480 h after LT exposure, respectively, and flash frozen in liquid nitrogen. Total RNA was isolated and the first-strand complementary DNA (cDNA) was synthesized. The qPCR primers were designed with Primer Premier 6.0 (http://www.premierbiosoft.com/primerdesign/index.html). The qRT-PCR was performed using SYBR Premix Ex Taq™ (Tli RNaseH Plus) (Takara #RR420Q, Takara Bio, Inc. China) in Roche Diagnostics with Light Cycler 480 System, and PCR amplification was conducted following the instructions. *Actin1* (*Csa3G806800*) was used as the reference gene for the normalization of gene expression values (Wan et al., 2010). Each experiment was conducted with three biological replicates and three technical replicates. The relative expression data of candidate gene was analyzed using $2^{-\Delta\Delta Ct}$ method (Kenneth et al., 2001).

Cloning and Sequence Analysis of Candidate Genes

The second true leaf of seedlings at the two-leaf stage was harvested for DNA extraction. Multiple pairs of primers were designed to amplify the full length candidate gene, and neighboring fragments have at least 100 bp overlap. Primers designed were listed in **Table S4**. PCR amplification was performed in a 25 µl reaction: 2.5 µl of 10X PCR buffer, 1 µl of 25X Mg^{2+}, 1 µl of 10X nucleoside triphosphate, 1 µl of 10 µM forward and reverse primer, 0.25 µl of KOD FX polymerase (Toyobo, Osaka, Japan) and 250–500 ng of DNA. The PCR conditions were as follows: 94°C for 5 min, 30 cycles of (94°C for 40 s, 55°C for 50 s, 72°C for 90 s), 72°C for 10 min, and a final incubation at 4°C for use. The PCR products were then sequenced by Sangon Biotech (Beijing, China). The sequence of the candidate genes of the parental lines were analyzed using SeqMan, and the amino acid sequences were analyzed using the MEGA 6.0 software (Tamura et al., 2013). The promoter sequences of candidate genes were analyzed using the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), and the

protein sequences of candidate genes were predicted by SMART software (Ponting et al., 1999).

Phylogenetic Analysis of *Csa6G445210* and *Csa6G445230*

To further understand the relatedness of the *Csa6G445210* and *Csa6G445230* sequences, phylogenetic trees were generated. The predicted amino acid sequences encoded by *Csa6G445210* and *Csa6G445230* in cucumber and orthologues in *Cucumis melon*, *Arabidopsis thaliana*, *Brassica rapa*, *Solanum lycopersicum*, *Oryza sativa* cv. *Japonica*, and *Zea mays* were obtained using the BLASTP tool (Altschul et al., 1997) in the NCBI database. Multiple sequence alignment was implemented with ClustalW (Thompson et al., 1994) and the phylogenetic tree was structured using the MEGA 6.0 software (Tamura et al., 2013) with a neighbor-joining algorithm.

Allelic Diversity of the Candidate Gene in F_2 Individuals

F_2 individuals were used to identify the correlation between the phenotype (*via* $F_{2:3}$ families) and the SNP mutation sites at the candidate genes. Specific SNP markers at mutation sites were designed (**Table S5**) to amplify DNA from F_2 individuals, and then the amplified products were digested for genotyping. The individuals that produce the same digestion products as "CG104," "CG37," and F_1 were defined as LT resistant (a), sensitive (b), and hybrid (ab), respectively. Then, the genotype and LTII value of each F_2 individual were compared.

Statistical Analysis

All tests for significant differences between "CG104" and "CG37" were done using one-way ANOVA in the R environment (R core Team, 2019).

RESULTS

Inheritance Analysis of Low-Temperature Tolerance in Cucumber Seedlings

To phenotype the LT tolerance of the two parental lines, and individuals in the F_1 and $F_{2:3}$ populations, 3-week-old seedlings were exposed to 15–17°C for 2 weeks in March 2017 and April 2017, respectively. The LTII was used to indicate the LT tolerance of each plant. Results showed that after 2 weeks of LT treatment, cotyledons and the first true leaf of "CG37" showed severe withering and yellowing, and the LTII score of "CG37" in greenhouse and plastic greenhouse were 59.1 and 57.7 respectively. However, "CG104" had no symptoms and the LTII were 14.9 and 8.7 respectively. The LTII of the F_1 hybrids were 26.4 and 21.7, respectively, with the symptoms more inclined toward "CG104." Frequency distribution of the LTII among $F_{2:3}$ families obeyed a normal distribution (**Figure 1** and **Figure S1**) in two experiments (the coefficient of correlation is 0.87). An additional experiment of phenotyping F_1 and F_1" showed that there is no

maternal effect (**Table S6**), which all together indicate that LT tolerance of cucumber at the seedling stage is a quantitative trait, controlled by multiple nuclear genes.

Linkage Map Construction and Quantitative Trait Locus Mapping

A total of 1,288 cucumber SSR markers were used to screen the polymorphisms between the parental lines; 509 of them showed polymorphisms, with a polymorphism rate of 39.5%. A total of 190 markers with clear electrophoresis strips were used to generate a linkage map (**Figure 2**). The markers selected were evenly distributed on seven chromosomes. The total length of the genetic map was 990.8 cM, and the average genetic distance between markers was 5.2 cM (**Table S7**). Based on the "9930" genome (Huang et al., 2009), the order of all markers on the genetic map was consistent with their physical location, therefore the map was qualified to be used for subsequent QTL mapping.

To detect QTL responsible for LT tolerance in cucumber seedlings, the phenotypic data for LT tolerance (LTII) from the two experiments and the genetic map constructed were used for QTL mapping. Details of each QTL detected, including map location, peak location, peak logarithm of odds (LOD) support value, confidence intervals, and percentages of total phenotypic variances explained (R^2) were shown in **Table 1**. In total, three QTLs including *qLTT5.1* on Chr.5, *qLTT6.1* and *qLTT6.2* on Chr.6 were repeatedly detected (with a threshold of 3, and the confidence intervals of 2-LOD) in two experiments (**Table 1** and **Figure 3**). Furthermore, LOD scores of *qLTT6.2* is 18.2 and 16.8, and accounts for 26.8 and 24.1% of the phenotypic variation. Therefore, it was proposed that *qLTT6.2* is a major QTL which is responsible for LT tolerance in cucumber seedlings.

Reducing the Interval of Locus *qLTT6.2* Using *In Silico* Bulked Segregant Analysis

Because *qLTT6.2* showed the highest contribution to LT tolerance in two experiments, we further conducted *in silico* BSA to narrow the region harboring this locus. Our previous study included the LT tolerance evaluation of 87 CG lines (**Figure S2**). Among these, five lines: "CG29," "CG56," "CG61," "CG90," "CG104" with strong LT resistance phenotypes, and another five lines: "CG10," "CG21," "CG43," "CG109," "CG37," with high LT sensitive phenotypes were used to generate LT-resistant and -sensitive bulks to identify the SNPs variation. The detailed information of these 10 lines including accession name, accession number in NCBI, geographical origin, and their LT tolerance performance are listed in **Table S1**. There were 214 non-synonymous SNPs that were identical *within* each bulk, but different *between* bulks (**Table S3**), all of which were distributed within the 42-kb region (20,779,616 to 20,821,620). These data indicated that this 42 kb region was associated with LTII variation among these lines. Annotation of the 42 kb genomic region predicted three genes including one for an auxin response factor, one for a CASP-like protein, and an *A. thaliana* ethylene insensitive 2 orthologue (**Figure 4** and **Table 2**).
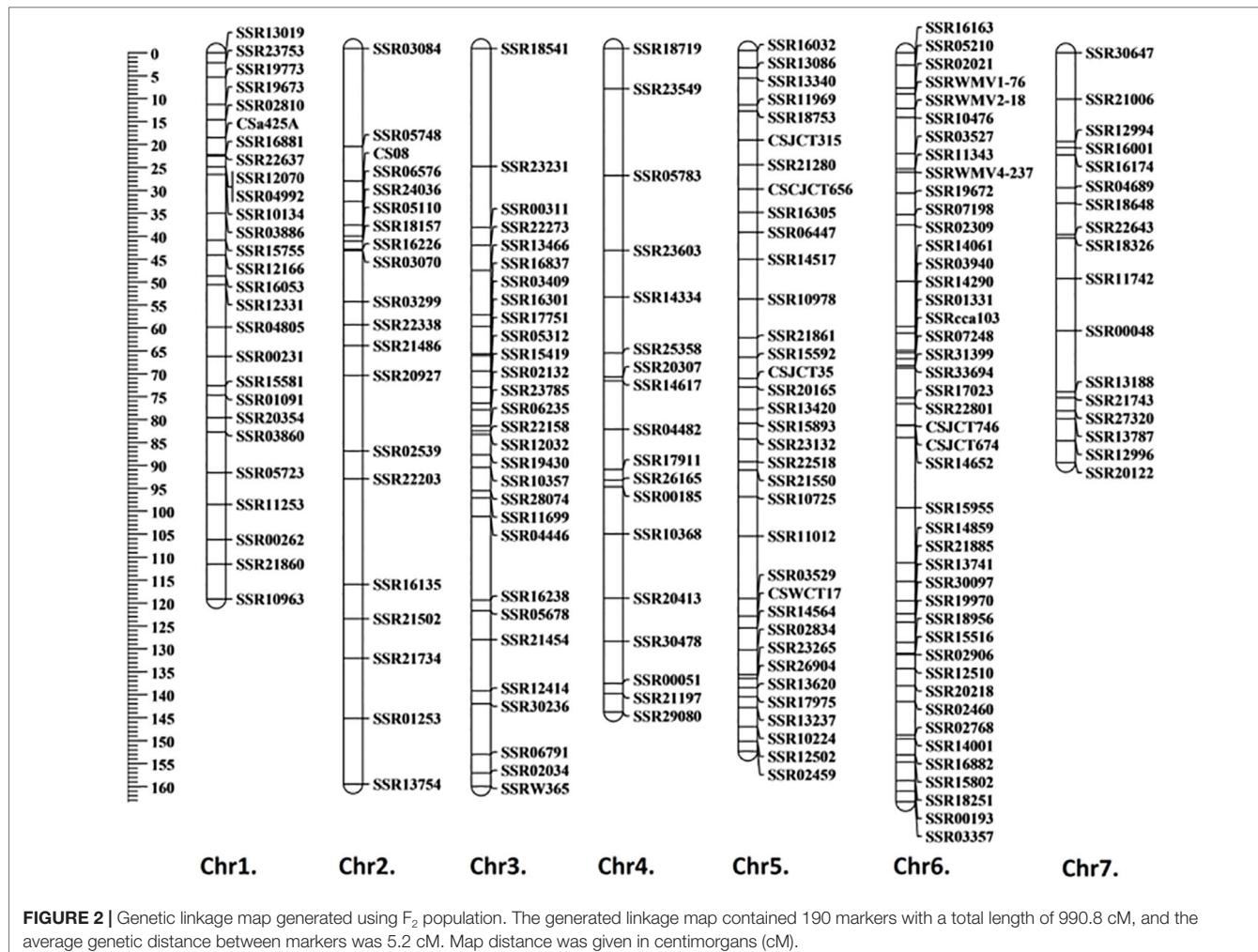
FIGURE 2 | Genetic linkage map generated using F_2 population. The generated linkage map contained 190 markers with a total length of 990.8 cM, and the average genetic distance between markers was 5.2 cM. Map distance was given in centimorgans (cM).

TABLE 1 | Identified quantitative trait loci controlling low temperature tolerance in cucumber seedlings. LTT indicates low temperature tolerance.

Treatments	QTL	Chr.	Peak location	Peak LOD	2 LOD interval (cM)	R^2/%
March 2017	qLTT1.1	1	18.5	3.6	SSR02810(14.6)–SSR16881(22.2)	8.1
	qLTT5.1	5	66.9	4.4	SSR15592(66.9)–CSJST35(71.5)	9.8
	qLTT6.1	6	63.1	6.4	SSR14290(61.1)–SSR01331(64.9)	8.0
	qLTT6.2	6	113.2	18.2	SSR21885(115.1)–SSR14859(119.3)	26.8
April 2017	qLTT5.1	5	66.9	6.0	SSR15592(66.9)–SSR20165(73.4)	13.5
	qLTT6.1	6	63.1	7.8	SSR14290(61.1)–SSR01331(64.9)	9.8
	qLTT6.2	6	113.2	16.8	SSR21885(115.1)–SSR14859(119.3)	24.1

Gene Expression Pattern Analysis of the Three Candidate Genes

To study the expression pattern of the three candidate genes under LT stress, leaf tissues of the two parental lines exposed to LT stress for 0, 10, 24, 48, 96, 144, 216, 288, 384, and 480 h were harvested, respectively, and then qRT-PCR of each gene was performed (**Figure 5**). The expression level of *Csa6G445210* in "CG104" was significantly higher than that of "CG37" at most time-points, and *Csa6G445230* showed a similar expression pattern. However, there was no significant expression difference

of *Csa6G445220* between the two parental lines. In short, *Csa6G445220* did not respond to LT stress, however, both *Csa6G445210* and *Csa6G445230* were up-regulated by LT stress in "CG104."

Cloning and Sequence Analysis of Candidate Genes

To further analyze the sequence of *Csa6G445210* and *Csa6G445230*, full-length DNA and cDNA of each gene in the two parental lines were sequenced and compared (**Figure 6**,

FIGURE 3 | Quantitative trait locus (QTL) analysis of low temperature tolerance in cucumber seedlings. Blue squares indicates the first treatment and red square indicates the second treatment. The x-axis indicates the genetic position of each chromosome, the y-axis indicates the logarithm of the odds value. Four QTLs were detected in two experiments and three QTLs were repeatedly identified.

FIGURE 4 | Identification of candidate genes for the *qLTT6.2* locus through combined linkage mapping and association analysis. **(A)** Linkage map for *qLTT6.2* locus on chromosome 6 based on 189 F$_{2:3}$ families from "CG37" × "CG104." **(B)** Distribution of 214 single-nucleotide polymorphisms in the "9930" cucumber genome, suggested that a 42-kb region was associated with the variation of low temperature tolerance in cucumber core germplasm materials. **(C)** Three candidate genes were identified in the 42-kb region.

TABLE 2 | Annotation of candidate genes responsible for low temperature tolerance.

Gene ID	Location on Chr. 6	Gene function annotation
Csa6G445210	20785976–20789697	Auxin response factor
Csa6G445220	20794324–20796912	CASP-like protein
Csa6G445230	20807729–20815381	Ethylene insensitive 2

Figures S3 and **S4**). The results showed that two base pair substitutions were detected within the coding sequence region (CDS) in *Csa6M445210*. The first substitution at position 1733 did not change the amino acid, however a polymorphism at position 1807 resulted in a non-synonymous change (Arg→Cys). Csa6M445230 has seven base pair substitutions within the CDS. Two substitutions were synonymous resulting in no amino acid change, while the rest were non-synonymous (Asp257→Ser, Thr820→Ser, Ser2468→Leu, Leu2678→Pro, Ser3260→Cys). The Asp257 resulted in changes in the transmembrane helix (**Table S8**).

The consistency between the LT tolerance phenotype, and the genotype of non-synonymous SNP sites within these two genes were examined in F_2 population. Primers were designed based on the non-synonymous SNP site of *Csa6G445210* and the first base substitution mutation site of *Csa6G445230*. By comparing the genotype and LTII value of each F_2 individual, it was found that both non-synonymous SNP sites within

Csa6G445210 and *Csa6G445230* were associated with the LT tolerance phenotypes (**Table S9**).

Phylogenetic Analysis of *Csa6G445210* and *Csa6G445230*

The gene annotation suggests that *Csa6M445210* encodes an auxin response factor (ARF), which is a transcription factor that controls the expression of auxin responsive genes. *Csa6M445230* encodes a transmembrane protein, ethylene-insensitive 2 (EIN2). To further predict and analyze the function of these two predicted amino acids, the sequences analyzed by BLAST against the NCBI database, and phylogenetic analyses were performed with their orthologues in seven other species (**Figure 7**). The results showed that the homologs of protein encoded by *Csa6M445210* and *Csa6G445230* are highly conserved in *A. thaliana* and other plant species, which suggests that they may share similar functions.

DISCUSSIONS

In our study, the LT stress treatments were carried out under greenhouse conditions in early spring, a natural LT environment in cucumber production. Using the yellowing degree of the first true leaf and the cotyledons as indicators, three QTLs were repeatedly identified on chromosome 5 and 6. The major QTL *qLTT6.2* was mapped into a 595-kb interval between markers SSR14859 and
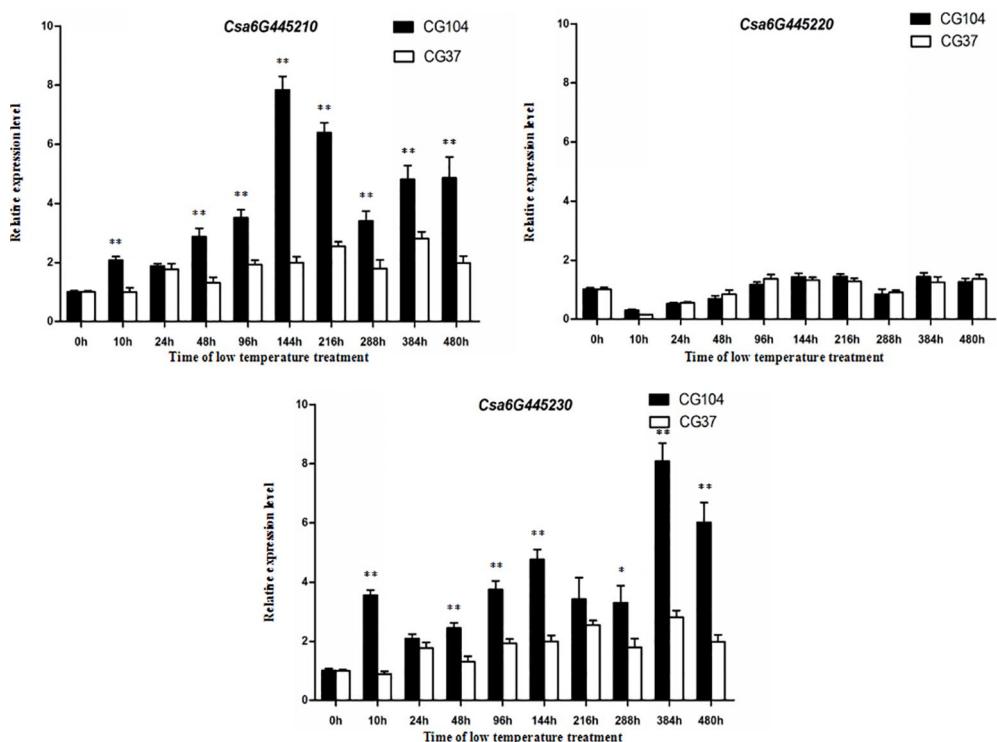
FIGURE 5 | Expression analysis of three candidate genes. The fold-change indicates the relative amount of gene transcripts at different time-points compared to 0 h. The asterisks indicate that there were significant differences in transcript level between "CG104" and "CG37" ("*," $0.01 < P < 0.05$; "**," $0.001 < P < 0.01$; "***," $0 < P < 0.001$).
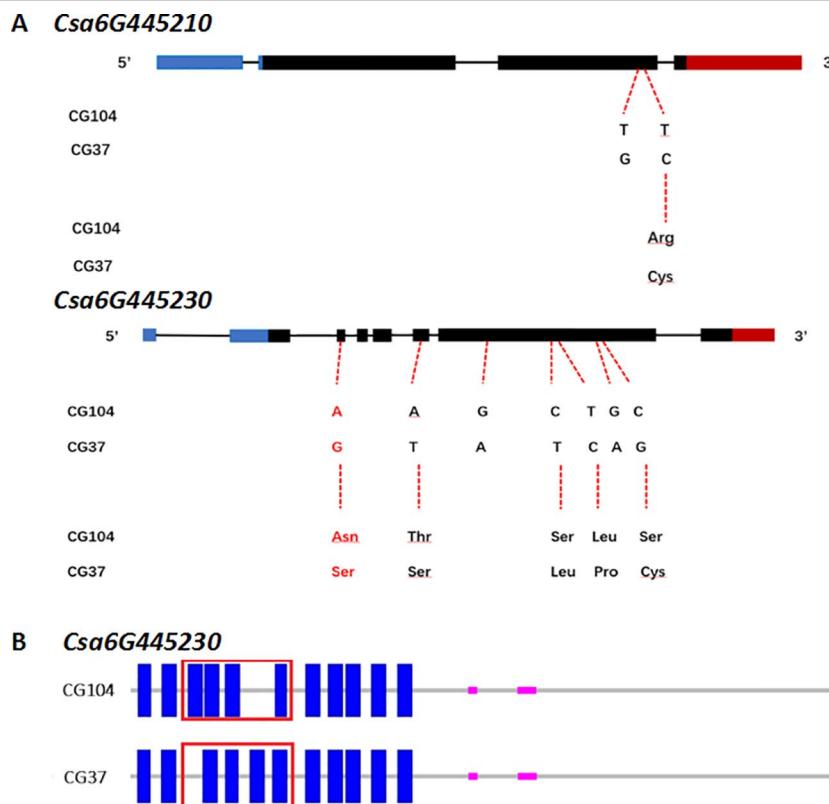
FIGURE 6 | Sequence analysis of *Csa6G445210* and *Csa6G445230*. **(A)** DNA and predicted amino acid changes between parental lines. Nucleic acid indicates the SNPs on exon of *Csa6G445210* and *Csa6G445230* between "CG104" and "CG37," and the consequential amino acid changes. The amino acid in red leads to the changes of protein secondary structure. **(B)** The secondary structure of protein encoded by *Csa6G445230* in "CG104" and "CG37." The blue solid rectangle indicates the transmembrane helix. The red box indicates the different transmembrane structure between "CG104" and "CG37."
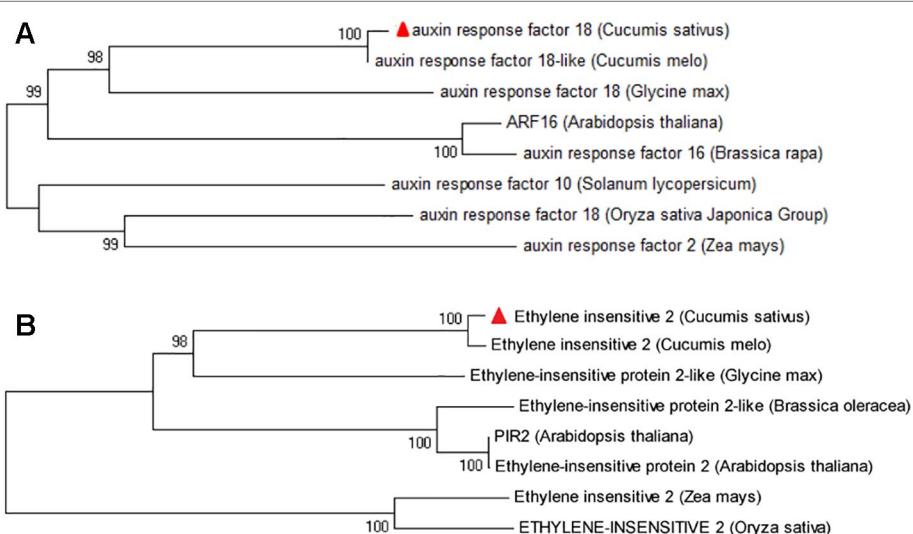
FIGURE 7 | Phylogenetic tree of protein encoded by *Csa6G445210* and *Csa6G445230*. The protein encoded by cucumber *Csa6G445210* **(A)** or *Csa6G445230* **(B)** is shown as a red triangle; the branches refer to the rates of amino acid variation.

SSR21885 with LODs support score of 18.2 and 16.8, and explained 26.8 and 24.1% of the observed phenotypic variations. Several studies on QTL mapping of LT-tolerance in cucumber seedlings have been reported. However, the results differed due to variation in the developmental stage of the plant materials used, differences in treatment conditions including location (environmental controlled facility or field) and temperature (12 or 4°C), and different evaluation standards (degree of yellowing, area of dehydration spot, or area of dryness after recovery). Li (2014) used a CII as indicator, identified a marker SSR07248 that was closely linked to a LT-tolerant locus on Chr.6. Zhou (2015) identified six QTLs associated wtih LT tolerance on chromosome 3, and Wang (2014) identified three QTLs on chromosome 3 and one QTL on chromosome 7 using CII and a recovery index as indicators of LT-stress. However, the flanking markers were not close enough, and the contribution rate of the QTLs was low in above studies. SSR01331, the flanking maker of $qLTT6.1$ identified in our study, was only 233 kb away from SSR07248 identified in the previous study (Li, 2014). The relationship between these two loci needs to be further investigated.

"In silico BSA" is an effective method to discover or refine markers linked to target genes or QTLs, by inspecting SNP variation among genotyped individuals at specific genomic regions. It has been used successfully in soybean to map virus resistant genes (Bello et al., 2014). With this method in cucumber, Li et al. (2016) quickly narrowed the interval around the gene controlling carpel number from a 1.9 Mb region down to 16 kb. Bo et al. (2019) also used this method to fine-map the fruit spine density related major QTL ($qfsd6.2$) to a 50 kb region, and finally identified the candidate gene $Csgl3$ for the trait. In our study, the phenotypic data obtained in two replicated experiments were consistent, and the genotype sequencing depth reached 30 X, therefore this method was employed to narrow the target region. Based on the sequence data and LT-tolerance performance of CG, the genomic region of $qLTT6.2$ was delimited to a 42-kb region, which only contained three candidate genes. Gene sequence alignments revealed that there were amino acid substitutions in both $Csa6G445210$ and $Csa6G445230$. Furthermore, allelic diversity of the candidate gene region in F_2 individuals showed that the two markers based on the non-synonymous SNP sites within $Csa6G445210$ and $Csa6G445230$ were both associated with the phenotypes of F_2 individuals. Therefore, our study demonstrated that the in silico BSA approach can accurately identify SNPs linked to LT tolerance trait in cucumber.

LT is one of the major abiotic stresses that affect plant growth, development, and productivity. The mechanisms plants have evolved to adapt to LT stress have been well-studied over the past two decades (Ding et al., 2019), however, little is known in cucumber. In our study, two candidate genes $Csa6G445210$ and $Csa6G445230$ that might be involved in LT-stress response in cucumber seedlings were identified. $Csa6G445230$ encodes EIN2 which is a key regulator in ethylene signaling pathway, and $Csa6G445210$ encodes ARF which is responsive to auxin. The role of ethylene in abiotic stress resistance including cold responses was established (Yoo et al., 2009; Shi et al., 2012). In Arabidopsis, EIN2 is a positive regulator of the ethylene signaling pathway, indirectly inhibiting the expression of CBFs, which is achieved by the downstream transcription factor EIN3/EILs (Guo and Ecker, 2003; An et al., 2010; Shi et al., 2012;

Li et al., 2015). CBFs could quickly activate the set of downstream cold-responsive genes COR (i.e., cold regulated genes), to improve plant cold tolerance (Gilmour et al., 1998; Medina et al., 1999; Miura and Furumoto, 2013; Barrero-Gil and Salinas, 2017; Shi et al., 2018). However, whether ethylene has a positive or negative effect on cold tolerance might vary among plant species; ethylene may positively regulate a low-temperature response in chilling-sensitive species (Eremina et al., 2016). In A. thaliana, ethylene content decreased when exposed to LT, and the loss-of-function mutant $ein2–5$ was more resistant to LT stress than the wild-type plants, which indicates that ethylene negatively regulates LT stress response in Arabidopsis (Shi et al., 2012). The same conclusions were drawn in other LT-tolerant plants such as Medicago sativa (Zhao et al., 2014) and Triticum aestivum (Field, 1984). However, in fruit of C. sativus (Wang and Adams, 1982) and S. lycopersicum (Zhao et al., 2009), studies showed that ethylene content increased during rewarming stage after LT treatment. Whether ethylene content increased during LT treatment in leaves is yet unknown. In addition, $CsEIN2$ was up-regulated in the LT-tolerant line in our study, we thus speculate that $CsEIN2$ might positively regulate CBF in cucumber seedlings, which deserves further investigation. Although auxin is a master regulator of plant growth and development, little is known regarding its role in cold stress response. Aux/IAA binds to ARF (auxin-responsive factor) and prevents the ARF-mediated expression of downstream genes (Lavy and Estelle, 2016). A recent study showed that several DREB/CBFs genes regulate the Aux/IAA genes directly, which demonstrate that the cold responsive pathway interacts with the auxin gene regulatory network (Shani et al., 2017). However, if and how $Csa6G445210$ regulates aspects of the cold tolerance signaling pathway should be determined.

DATA AVAILABILITY STATEMENT

The accession numbers of the ten core germplasm lines used in our study can be found in **Table S1**. The link of the sequence data in NCBI Short Read Archive (SRA) SRA056480 is as follows: (https://www.ncbi.nlm.nih.gov/sra/?term=SRA056480). All SNPs within the 42 kb target region can be found in **Table S3**.

AUTHOR CONTRIBUTIONS

WW and SW conducted the experiments and analyzed the data. SD, WW and KB analyzed the data and drafted the manuscript. ZS helped collect the data. HM helped analyze the data. XG and SZ designed the experiments.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01620/full#supplementary-material

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25 (17), 3389–3402. doi: 10.1093/nar/25.173389

An, F., Zhao, Q., Ji, Y., Li, W., Jiang, Z., Yu, X., et al. (2010). Ethylene-induced stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 is mediated by proteasomal degradation of EIN3 binding F-box 1 and 2 that requires EIN2 in Arabidopsis. *Plant Cell* 22 (7), 2384–2401. doi: 10.1105/tpc.110.076588

Barrero-Gil, J., and Salinas, J. (2017). CBFs at the crossroads of plant hormone signaling in cold stress response. *Mol. Plant* 10 (4), 542–544. doi: 10.1016/j.molp.2017.03.004

Bello, M. H., Moghaddam, S. M., Massoudi, M., McClean, P. E., Cregan, P. B., and Miklas, P. N. (2014). Application of *in silico* bulked segregant analysis for rapid development of markers linked to bean common mosaic virus resistance in common bean. *BMC Genomics* 15, 903. doi: 10.1186/1471-2164-15-903

Bo, K., Miao, H., Wang, M., Xie, X., Song, Z., Xie, Q., et al. (2019). Novel loci fsd6.1 and Csgl3 regulate ultra-high fruit spine density in cucumber. *Theor. Appl. Genet.* 132 (1), 27–40. doi: 10.1007/s00122-018-3191-6

Cabrera, R. M., Saltveit, M. E.Jr., and Owens, K. (1992). Cucumber cultivars differ in their response to chilling temperatures. *J. Am. Soc. Hortic. Sci.* 117, 802–807. doi: 10.21273/JASHS.117.5.802

Chen, B., Saltveit, M. E., and Beckles, D. M. (2019). Chilling-stress modifies DNA methylation level in cucumber (*Cucumis sativus* L.) seedling radicle to regulate elongate rate. *Sci. Hortic.* 252, 14–19. doi: 10.1016/j.scienta.2019.03.023

Chung, S. M., Staub, J. E., and Fazio, G. (2003). Inheritance of chilling injury: a maternally inherited trait in cucumber. *J. Am. Soc. Hort. Sci.* 128, 526–530. doi: 10.1111/nph.15696

Ding, Y., Shi, Y., and Yang, S. (2019). Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. *New Phytol.* 222 (4), 1690–1704. doi: 10.1007/s00018-015-2089-6

Eremina, M., Rozhon, W., and Poppenberger, B. (2016). Hormonal control of cold stress responses in plants. *Cell Mol. Life Sci.* 73 (4), 797–810. doi: 10.1093/oxfordjournals.aob.a086774

Field, R. J. (1984). The role of 1-aminocyclopropane-1-carboxylic acid in the control of low temperature induced ethylene production in leaf tissue of *Phaseolus vulgaris* L. *Ann. Bot.* 54, 61–67. doi: 10.1046/j.1365-313x.1998.00310.x

Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M., et al. (1998). Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J.* 16, 433–442. doi: 10.1016/S0092-8674(03)00969-3

Guo, H., and Ecker, J. R. (2003). Plant responses to ethylene gas are mediated by SCF(EBF1/EBF2)-dependent proteolysis of EIN3 transcription factor. *Cell* 115, 667–677. doi: 10.1006/meth.20011262

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

Kenneth, J. L., Thomas, D. S., 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: 10.21273/JASHS.133.2.225

Kozik, E. U., and Wehner, T. C. (2008). A single dominant gene Ch for chilling resistance in cucumber seedlings. *J. Am. Soc. Hort. Sci.* 133, 225–227. doi: 10.1093/nar/gkq1019

Kozik, E. U., Klosinska, U., and Wehner, T. C. (2012). Inheritance of low temperature seed germination ability in cucumber. In: Proc. of 10th Eucarpia International Meeting on Genetics and Breeding Cucurbitaceae, 575–578. doi: 10.1242/dev.131870.

Lavy, M., and Estelle, M. (2016). Mechanisms of auxin signaling. *Development* 143 (18), 3226–3229. doi: 10.1007/s00122-016-2679-1

Leionen, R., Sugawara, H., and Shumway., M. (2011). International nucleotide sequence database collaboration. The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1016/j.cell.2015.09.037

Li, W., Ma, M., Feng, Y., Li, H., Wang, Y., Ma, Y., et al. (2015). EIN2-directed translational regulation of ethylene signaling in arabidopsis. *Cell* 163, 670–683. doi: 10.1104/pp.119.2.463

Li, S., Pan, Y., Wen, C., Li, Y., Liu, X., Zhang, X., et al. (2016). Integrated analysis in bi-parental and natural populations reveals CsCLAVATA3 (CsCLV3) underlying carpel number variations in cucumber. *Theor. Appl. Genet.* 129, 1007–1022. doi: 10.1007/s00122-016-2679-1

Li, H. (2014). Genetic analysis and linked marker selection of low temperature tolerance in cucumber. [master's thesis]. Shanghai: Shanghai JiaoTong University. doi: 10.3390/ijms14035312.

Medina, J., Bargues, M., Terol, J., Pérez-Alonso, M., and Salinas, J. (1999). The Arabidopsis CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiol.* 119, 463–469. doi: 10.1093/nar/27.1.229

Miura, K., and Furumoto, T. (2013). Cold signaling and cold response in plants. *Int. J. Mol. Sci.* 14, 5312–5337. doi: 10.1038/ng2801

Ponting, C. P., Schultz, J., Milpetz, F., and Bork, P. (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res.* 27 (1), 229–232. doi: 10.1371/journal.pone.0005795

Qi, J. J., Liu, X., Shen, D., Miao, H., Xie, B., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45, 1510–1515. doi: 10.1016/j.cub.2016.12.016

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. (Vienna, Austria). https://www.R-project.org/.

Ren, Y., Zhang, Z., Liu, J., Staub, J. E., Han, Y., et al. (2009). An integrated genetic and cytogenetic map of the cucumber genome. *PloS One* 4, e5795. doi: 10.1093/nar/22.224673

Shani, E., Salehin, M., Zhang, Y., Sanchez, S. E., Doherty, C., Wang, R., et al. (2017). Plant stress tolerance requires auxin-sensitive Aux/IAA transcriptional repressors. *Curr. Biol.* 27 (3), 437–444. doi: 10.1016/j.tplants.2018.04.002

Shi, Y., Tian, S., Hou, L., Huang, X., Zhang, X., Guo, H., et al. (2012). Ethylene signaling negatively regulates freezing tolerance by repressing expression of CBF and Type-A ARR Genes in Arabidopsis. *Plant Cell.* 24, 2578–2595. doi: 10.1371/journal.pone.0167845

Shi, Y., Ding, Y., and Yang, S. (2018). Molecular regulation of CBF signaling in cold acclimation. *Trends Plant Sci.* 23 (7), 623–637. doi: 10.1023/A:1003084821178

Smeets, L., and Wehner, T. C. (1997). Environmental effects on genetic variation of chilling resistance in cucumber. *Euphytica* 97, 217–225. doi: 10.1093/molbev/mst197

Song, Z. C., Miao, H., Zhang, S., Wang, Y., Zhang, S. P., and Gu, X. F. (2016). Genetic analysis and QTL mapping of fruit peduncle length in cucumber (Cucumis sativus L.). *PloS One* 11, e0167845. doi: 10.1007/BF00227388

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30 (12), 2725–2729. doi: 10.1016/j.ab.2009.12.008

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi: 10.1105/tpc.112.098640

Wan, H., Zhao, Z., Qian, C., Sui, Y., Malik, A. A., and Chen, J. (2010). Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. *Anal. Biochem.* 399, 257–261. doi: 10.1016/j.ab.2009.12.008

Wang, C. Y., and Adams, D. O. (1982). Chilling-induced ethylene production in cucumbers (*Cucumis sativus* L.). *Plant Physiol.* 69, 424–427.

Wang, H. J., Wu, Y., Gu, W., Sun, X. D., and Qin, Z. W. (2006). Extraction of DNA from cucumber by improved CTAB method. *Heilongjiang Agric. Sci.* 5, 124–125.

Wang, W. P., Song Z.C., B. O. K. L., Dong, S. Y., Wei, S., Miao, H., Li, J. B., et al. (2019). Evaluation and genome-wide association study (GWAS) of low-temperature tolerance at seedling stage in cucumber core germplasm. *J. Plant Genet. Resour.* 20 (6), 1606–1612. doi: 10.13430/j.cnki.jpgr.20190225001

Wang, H. (2014). Evaluation of low temperature tolerance and QTL mapping in cucumber germplasm. [Doctoral dissertation]. Beijing: Chinese Academy of Agriculture Sciences. doi: 10.1007/s00122-015-2544-7.

Weng, Y., Colle, M., Wang, Y., Yang, L., Rubinstein, M., Sherman, A., et al. (2015). QTL mapping in multiple populations and development stages reveals dynamic QTL for fruit size in cucumbers of different market classes. *Theor. Appl. Genet.* 128, 1747–1763. doi: 10.1016/j.febslet.2009.09.029

Yoo, S. D., Cho, Y., and Sheen, J. (2009). Emerging connections in the ethylene signaling network. *Trends Plant Sci.* 14, 270–279. doi: 10.1111/ppl.12161

Zhao, D., Shen, L., Fan, B., Yu, M., Zheng, Y., et al. (2009). Ethylene and cold participate in the regulation of *LeCBF1* gene expression in postharvest tomato fruits. *FEBS Lett.* 583, 3329–3334. doi: 10.1016/j.febslet.2009.09.029

Zhao, M., Liu, W., Xia, X., Wang, T., and Zhang, W. H. (2014). Cold acclimation-induced freezing tolerance of Medicago truncatula seedlings is negatively regulated by ethylene. *Physiol. Plant* 152, 115–129. doi: 10.1111/ppl.12161

Zhou, S. (2015). Evaluation of low temperature tolerance and molecular marker in cucumber germplasm. [master's thesis]. Changchun: Northeast Agriculture University.

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.

Molecular Genetic Mapping of Two Complementary Genes Underpinning Fruit Bitterness in the Bottle Gourd (*Lagenaria siceraria* [Mol.] Standl.)

Xiaohua Wu†, Xinyi Wu†, Ying Wang, Baogen Wang, Zhongfu Lu, Pei Xu and Guojing Li*

Institute of Vegetables, State Key Laboratory for Quality and Safety of Agro-products, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

Fruit bitterness is a serious problem threatening the bottle gourd (*Lagenaria siceraria* [Mol.] Standl.) industry worldwide. Previous genetic studies indicated that fruit bitterness in the bottle gourd was controlled by a pair of complementary genes. In this study, based on two non-bitter landraces "Hangzhou Gourd" and "Puxian Gourd," each of which carries a single bitterness gene, and their derived segregation populations, we mapped the complementary genes causing fruit bitterness. Quantitative trait locus (QTL) scanning based on an F_2 population detected two QTLs, which was *QBt.1* locating in a 17.62-cM interval on linkage group (LG)2 corresponding to a 1.6-Mb region on chromosome 6, and *QBt.2* mapped to a 8.44-cM interval on LG9 corresponding to a 1.9-Mb region on chromosome 7. An advanced bulked segregant analysis (A-BSA) well validated the QTL mapping results. Sequence-based comparative analysis showed no syntenic relationship between *QBt.1/QBt.2* and the known bitterness genes in cucumber, melon, and watermelon, suggesting that causal genes underlying *QBt.1* and *QBt.2* were not direct orthologs of the reported cucurbit bitterness genes. Our results shed light on the molecular genetic mechanisms underlying fruit bitterness in the bottle gourd and is useful to guide breeders to properly select parental lines to avoid the occurrence of bitter fruits in breeding programs.

Keywords: bottle gourd, fruit bitterness, complementary genes, genetic mapping, comparative analysis

INTRODUCTION

Bottle gourd or calabash (*Lagenaria siceraria* [Mol.] Standl.) ($2n = 2x = 22$), a member of the genus *Lagenaria* of the Cucurbitaceae family (Beevy and Kuriachan, 1996), is recognized as indigenous to Africa and domesticated independently in Asia (Erickson et al., 2005). Having a cultivation history of over 8,000 years by man (Crawford, 1992), today, the bottle gourd is grown all over the tropics and subtropics for its immature fruits used as a vegetable or hard-shelled mature fruits used as containers, musical instruments, or handicrafts. The bottle gourd seedlings are widely used as a rootstock for grafting with watermelon to defend soil-borne diseases and to increase low-temperature tolerance (Davis et al., 2008; King et al., 2008).

While young fruits of bottle gourd are traditionally consumed as a delightful culinary vegetable in many areas of Asia including China, the undesirable occurrence of fruit bitterness in this crop as documented in the ancient medical books *Ben Cao Jing Ji Zhu* and *Ben Cao Gang Mu* has long been

threatening the bottle gourd industry. Fruit bitterness not only affects the economic value of the bottle gourd but causes severe food poisoning symptoms such as nausea, vomiting, diarrhea, and abdominal cramps in humans and livestock (Zhang, 1981). Fruit bitterness appears to be a common trait to the Cucurbitaceae family, in which the compounds cucurbitacins having the function of defending against insects and herbivores cause the bitterness phenotype (Balkema-Boomstra et al., 2003). In cucumber, the effective cucurbitacin component is cucurbitacin C (CuC). It has been clear that nine genes are involved in the CuC biosynthetic pathway, and two genes, Bl (Bitter leaf) and Bt (Bitter fruit), regulate the bitterness/non-bitterness phenotype in leaves and fruits, respectively. The selection of natural mutations on Bi and Bt has played an important role in the domestication of ancient wild cucumber to form the present-day non-bitter cultivars (Shang et al., 2014). The major gene clusters for cucurbitacin biosynthesis were found to be highly conserved in cucumber, melon, and watermelon, while the regulatory genes seemed divergent among the three crops (Zhou et al., 2016).

Previous studies through traditional genetic analysis indicated that fruit bitterness in the bottle gourd was controlled by a pair of complementary genes (Zhang, 1981); however, the genome locations of the genes remain unknown. Genome-wide scan for quantitative trait loci (QTLs) is an effective approach for mapping traits governed by multi-genes including digenic genes (Zhang et al., 2013). Alternatively, Wu and Huang (2006) developed an advanced bulked segregant analysis (A-BSA hereafter) method to specifically identify DNA markers linked to two interactive genes. This method relies on the construction of a DNA pool mixed with homozygous genotypes (homo-pool) and another pool mixed from individuals with heterozygous genotypes showing contrasting phenotypes (heter-pool). By comparing marker genotypes among the parental lines and the two pools, DNA markers linked to the trait-determining genes as well as their parental allele origins can be inferred.

In the current study, we mapped the fruits bitterness genes in the bottle gourd genome and revealed their relationships with known bitterness genes in related food cucurbits. Our study also demonstrates the usefulness of combining QTL scanning and A-BSA as a fast and efficient approach to map and validate complementary genes controlling a trait.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Two bottle gourd landraces "Hangzhou Gourd" (HZ hereafter) and "Puxian Gourd" (PX hereafter), their F_1 progenies, and F_2 populations derived from selfing independent F_1s in the years 2013, 2014, and 2016 were used in this study. The population size for the three F_2 populations (2013F_2, 2014F_2, and 2016F_2) were 102, 169, and 101, respectively. For A-BSA, 24 individuals showing a bitter-fruit phenotype from the 2013F_2 population were screened for their F_3 progenies phenotypes, and those showing no fruit bitterness segregation in F_3 generation were selected to construct the homozygous bitter pool (homo-pool). Six F_2 individuals with non-bitter fruits were randomly selected to construct a heter-pool (**Figure 1**). All plants were grown in tunnel houses in 30-m rows spaced 0.5 m apart. Ambient temperature and light as well as normal management were applied.
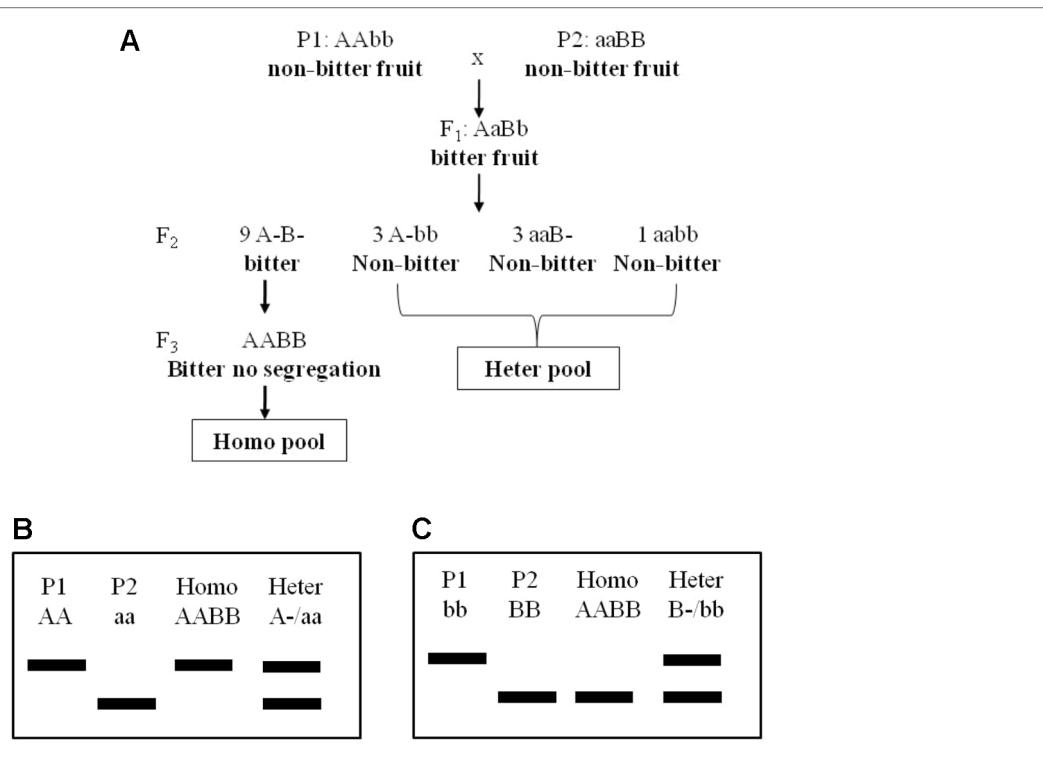
FIGURE 1 | The genetic model of fruit bitterness in the F_2 population **(A)** and advanced BSA method using in this study **(B, C)**. BSA, bulked segregant analysis.

Fruit Bitterness Evaluation

Fruit bitterness evaluation was conducted at the stage of 8–12 days after pollination. Bitterness phenotype of each individual was determined by manually tasting the sliced fruit sarcocarp by three trained tasters according to a classic tasting method described by Andeweg and DeBruyn (1959). The bitterness phenotype was scored as 1 for bitter and 0 for non-bitter. Only those results that were consistent among all three tasters were considered trustable and used in further analysis.

DNA Extraction and Single-Nucleotide Polymorphism Genotyping

Genomic DNA was extracted from young leaves of 2-week-old seedlings using a DNA extraction kit (TIANGEN Co. Ltd, Beijing) following the manufacturer's instructions. For single-nucleotide polymorphism (SNP) genotyping of the mapping populations, our previous RAD-Seq data from a set of bottle gourd accessions including HZ and PX were revisited (Wu et al., 2017a), from which 192 out of the 684 SNPs between the two parents that were evenly distributed in the genome were selected. Kompetitive allele-specific PCR (KASP) assays were used to genotype the mapping populations. KASP primers were designed using the Kraken™ software system (https://www.biosearchtech.com/support/tools/genotyping-software/kraken). Each KASP reaction was carried out in a final volume of 10 μl containing 20–40 ng of genomic DNA, 5 μl of 2× premade KASP master mix (LGC, Middlesex, UK), and 0.14 μl of primer mix. PCR amplification was performed in a Hydrocycler[2] water bath thermal cycler following LGC parameters: 95°C, 15 min for hot-start Taq DNA polymerase activation, followed by a touchdown profile of 10 cycles at 94°C for 20 s and 61°C for 1 min with a 0.6°C reduction per cycle, and followed by 26 cycles at 94°C for 20 s and 55°C for 1 min. End-point fluorescent images were visualized using the BMG FLUOstar Omega (https://www.biosearchtech.com/products/instruments-and-consumables/genotyping-instruments/snpline-genotyping-automation/plate-reading), and allele calls for each genotype were obtained using the KlusterCaller™ software (LGC, UK).

Linkage Mapping and Quantitative Trait Locus Analysis

A genetic linkage map for SNPs was constructed based on the 2014F_2 population using the software QTL IciMapping (http://www.isbreeding.net). A likelihood of odds (LOD) threshold of 3.0 was used to determine the linkage groups (LGs), and an nnTwoOpt algorithm was used to determine the maker orders in each LG. The software MapQTL V5 (Van Ooijen, 2004) was used to detect bitterness QTL. Firstly, the interval mapping (IM) model was applied to detect QTLs for fruit bitterness, and then a multiple-QTL model (MQM) was used to scan for new QTLs with the markers closest to the original QTLs being implemented as cofactors. The mapping parameters were as follows: step size = 1.0, maximum number of neighboring markers = 5, maximum number of iterations = 200, and function tolerance = $1.0e^{-08}$. A genome-wide permutation test with 1,000 random permutations was conducted to obtain an empirical LOD score threshold for significance ($P < 0.05$).

Comparative Mapping of Bitterness Genes

The coding DNA sequence (CDS) of the bitterness genes *Bt*, *Bi*, and *Bl* from cucumber, melon, and watermelon (Shang et al., 2014; Zhou et al., 2016) were Blastn searched against the latest Hangzhou Gourd reference genome assembly V2.0 (Wang et al., 2018) to locate their syntenic region and orthologous genes in the bottle gourd genome with an *e*-value cutoff of 1*e*−10. The annotation of genes in the bottle gourd QTL intervals was retrieved from GourdBase (Wang et al., 2018).

RESULTS

Inheritance of Fruit Bitterness in the Mapping Populations

Fruit bitterness assessment of parental lines, their F_1 progenies, and F_2 populations showed consistent phenotypes in the three biological replicates in 2013, 2014, and 2016. HZ and PX always showed a non-bitter phenotype, whereas their F_1 progenies were always bitter. For each of the three F_2 sub-populations, the bitter- and non-bitter-fruited individuals fit a 9:7 segregating ratio (**Table 1**), which suggested two complementary genes controlling this trait as previously reported in other cultivars of this species (Zhang, 1981).

Quantitative Trait Locus Mapping for Fruit Bitterness

Of the 192 intended KASP assays, 174 successfully detected signals and 153 called polymorphic genotypes in the F_2 population in 2014. Based on these 153 SNPs data, a genetic linkage map containing 147 SNP loci distributed on 11 LGs were constructed. In line with the digenic mode of fruit bitterness inheritance in this population, two QTLs, designated as *QBt.1* and *QBt.2*, were detected through IM followed by MQM (**Table 2, Figure 2** and **Supplementary Table 1**). *QBt.1* was located in a 17.62-cM interval between the SNP markers BGReSe_09031 and BGReSe_09068 on LG2, which explained 18% of the phenotype variance; *QBt.2*

TABLE 1 | Fruit bitterness phenotypes.

Genotype/population	Number of individuals			
	Bitter	**Non-bitter**	**χ^2**	**Ratio**
Hangzhou Gourd	0	48		
Puxian Gourd	0	48		
F_1	48	0		
2013F_2	59	43	0.16	9:7
2014F_2	106	63	2.65	9:7
2016F_2	63	38	1.37	9:7

TABLE 2 | QTLs for fruit bitterness detected in the 2014 F_2 population.

QTL	LG	Flanking marker	Genetic position	LOD	R^2 (%)
QBt.1	LG2	BGReSe_09031-BGReSe_09068	0–17.62	8.35	18
QBt.2	LG9	BGReSe_11107-BGReSe_11032	50.37–58.81	12.86	27.7

QTL, quantitative trait locus; LG, linkage group.

FIGURE 2 | QTLs detected in the F$_2$ population under a multiple-QTL model using MAPQTL V5. QTL, quantitative trait locus.
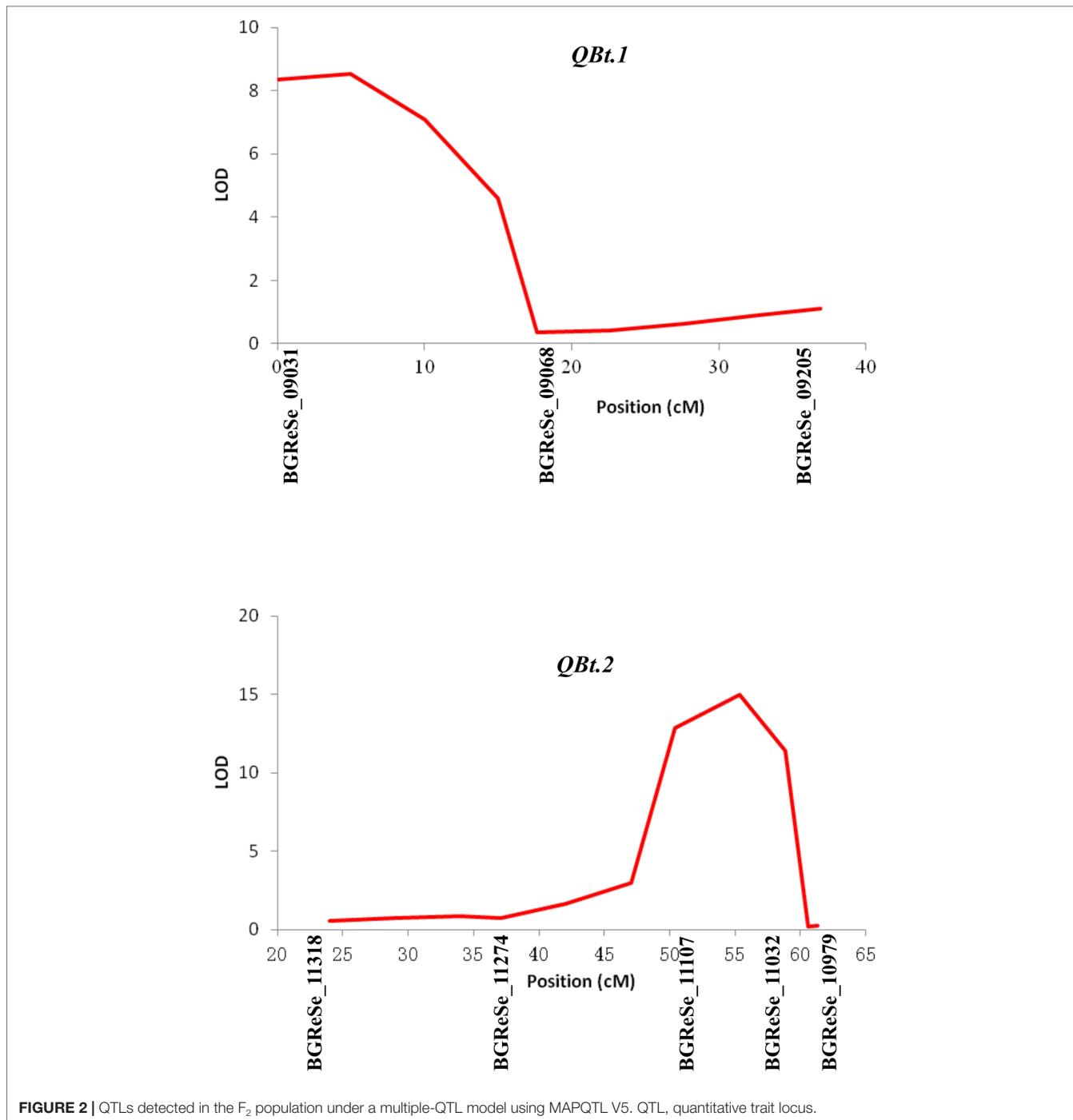
was mapped to an 8.44-cM interval defined by the SNP markers BGReSe_11107 and BGReSe_11032 on LG9, which accounted for 27.7% of the phenotype variance. QTL epistatic interaction was detected between *QBt. 1* and *QBt.2*, and the phenotypic variation explained reached 99.32%, indicating that *QBt.1* and *QBt.2* were the pair of complementary genes controlling fruit bitterness in this population.

TABLE 3 | The genotypes of linked markers detected using the advanced BSA method.

SNP	Hangzhou Gourd	Puxian Gourd	Homo-pool	Heter-pool
BGReSe_09068	TT	CC	TT	TC
BGReSe_11107	TT	CC	CC	TC

BSA, bulked segregant analysis; SNP, single-nucleotide polymorphism.
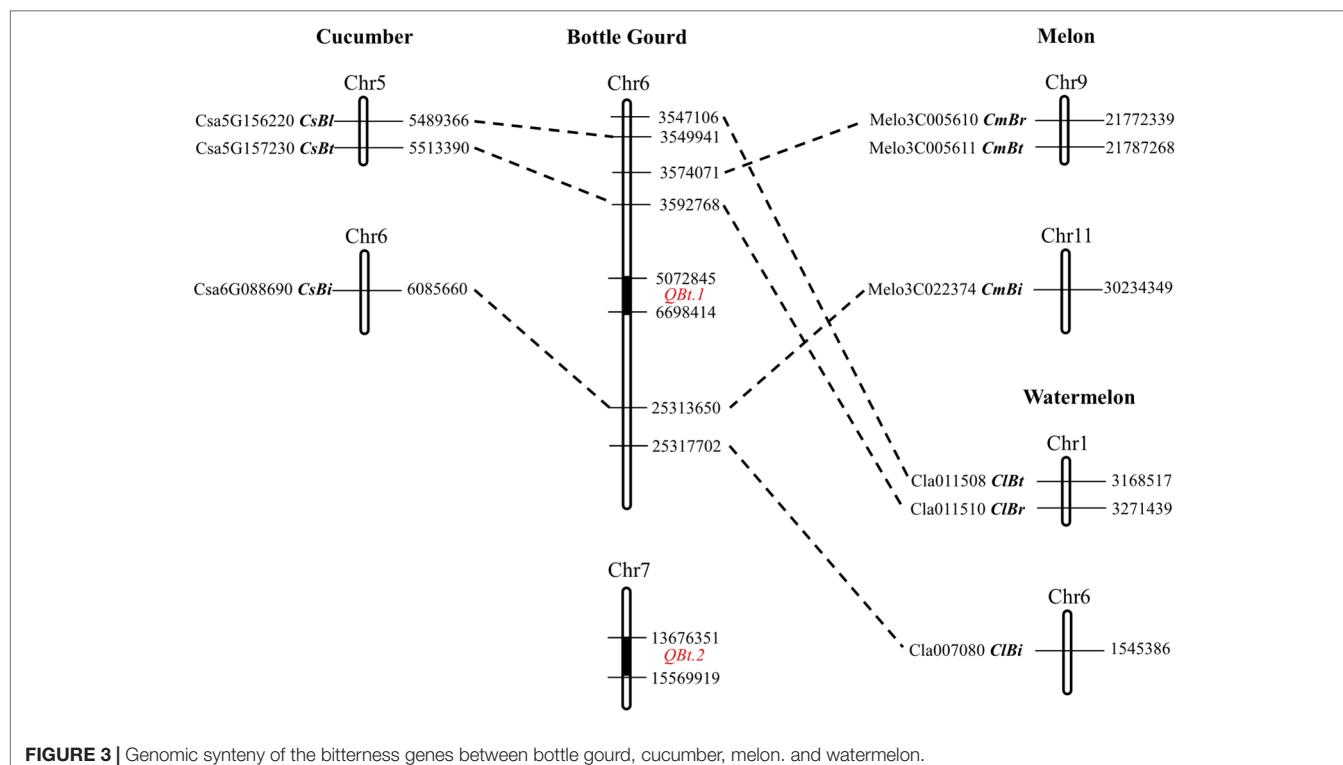
FIGURE 3 | Genomic synteny of the bitterness genes between bottle gourd, cucumber, melon. and watermelon.

Advanced Bulked Segregant Analysis for Single-Nucleotide Polymorphisms Linked to Fruit Bitterness

Another mapping approach, A-BSA, was also applied to screen for DNA markers linked to fruit bitterness. We screed 192 genome-wide distributed SNPs among the two parental lines, the homo-pool and the heter-pool, by using the KASP technology. It turned out that the marker BGReSe_09068 showed identical homozygous genotypes between HZ and the homo-pool, while it exhibited an alternative homozygous genotype in PX and, as expected, heterozygous genotype in the heter-pool (**Table 3**). Likewise, the marker BGReSe_11107 showed identical homozygous genotype in PX and the homo-pool but alternative homozygous genotype in HZ and heterozygous genotype in the heter-pool (**Table 3**). These results suggested that BGReSe_09068 and BGReSe_11107 each were linked to one of the complementary genes in HZ and PX, respectively. Coincidently, these two markers were found to fall into the QTLs intervals of *QBt.1* and *QBt.2*, respectively. These results thus provide strong validation for the QTL mapping result and indicates that *QBt.1* is one of the bitterness gene alleles from HZ and *QBt.2* is the other from PX.

Comparative Analysis of Bitterness Loci in Major Cucurbit Crops

According to the physical locations of the QTL-flanking markers on the HZ reference genome V2.0, *QBt.1* resides in a 1.62-Mb region (5,072,845–6,698,414 bp) on chromosome 6 with 142 predicted genes, and *QBt.2* in a 1.89-Mb region (13,676,351–15,569,919 bp) containing 89 predicted genes on chromosome 7. To elucidate the relationship between the bottle gourd fruit bitterness QTLs and the known bitterness genes in related cucurbits, a cross-mapping analysis was performed. It showed that the cucumber bitterness genes *CsBt* and *CsBi* and the *Bi* regulator gene *CsBl* were syntenic to the genomic region on chromosome 6 in the bottle gourd (**Figure 3**, **Supplementary Table 2**). The four tissue-specific cucurbitacin regulator genes (*CmBt* and *CmBr*, *ClBt* and *ClBr*) in melon and watermelon were also found to be syntenic to the same region on chromosome 6 in bottle gourd except for *CmBt*. *CmBi* in melon and *ClBi* in watermelon found their orthologous genes in the distal region on chromosome 6 in bottle gourd (**Figure 3**, **Supplementary Table 2**). However, neither *QBt.1* nor *QBt.2* was located in or close to the syntenic region of these known bitterness genes. In addition, all the cucurbitacin biosynthetic genes in cucumber, melon, and watermelon found orthologs in the bottle gourd genome, but none of them were located in the QTL regions (**Supplementary Table 2**). These results suggest that the causal genes underlying *QBt.1* and *QBt.2* are unlikely to be direct orthologs of the bitterness genes in these related cucurbit crops.

DISCUSSION

It has long been known that fruit bitterness in the bottle gourd is controlled by two interacting genes with complementary effect (Zhang, 1981). However, due to the lack of genomic resources, genetic mapping and elucidation of the bitterness genes in this species lag behind. Recently, plenty of SSR, Indel, and SNP markers information for the bottle gourd were released (Xu et al., 2011; Wu et al., 2017b; Wang et al., 2018), and reference genomes were available to the

public (Xu et al., 2014; Wu et al., 2017c; Wang et al., 2018), allowing for more in-depth characterization of bitterness genes in this species. In this study, we initially mapped two QTLs, *QBt.1* and *QBt.2*, each to a less than 2-Mb segment on chromosome 6 and chromosome 7, respectively. Then, by applying an advanced BSA method, we validated the QTL mapping results and elucidated that the functional alleles of *QBt.1* and *QBt.2* were from HZ and PX, respectively.

The ancestor wild cucumber, melon, watermelon, and bottle gourd plants all exhibited an extreme bitter phenotype in fruits (Zhou et al., 2016). Common bitterness compounds and their biosynthetic pathways were found in these related cucurbit crops (Lester, 1997; Matsuo et al., 1999; Chen et al., 2005). Likely due to common human demands for fruit quality, the domestication roadmap from extremely bitter wild ancestor to non-bitter modern cultivars seems to be also similar among cucumber, melon, and watermelon (Zhou et al., 2016). It has been argued that the causative mutations at orthologous genes may underlie convergent changes in fruit bitterness in cucurbits. For example, the loss of fruit bitterness during domestication in cucumber, melon, and watermelon all were caused by mutations on *Bt* genes (*CsBt*, *ClBt*, and *CmBt*) that are syntenic between genomes (Zhou et al., 2016). The cucumber *CsBt* gene encodes a basic helix–loop–helix (bHLH) transcription factor that activates the foliar bitterness gene *Bi* and regulates CuC biosynthesis in the fruit (Shang et al., 2014). According to the HZ reference genome V2.0, a gene *HG_GLEAN_10009202* was found to be the putative orthologous gene of *Bt*; however, this gene does not fall into the QTL intervals of *QBt.1* nor *QBt.2*. In addition, the bottle gourd orthologs of the cucumber foliar bitterness genes *CsBi* and *CsBl*, and the melon and watermelon root bitterness genes *CmBr* and *ClBr*, were all found to be not coincident with *QBt.1* or *QBt.2*. Therefore, the causal genes underlying the bitterness QTLs in the bottle gourd are unlikely to be direct orthologs of the cucumber *CsBt*, *CmBr*, or *ClBr* gene. However, this result does not exclude the possibility that similar genes or same gene family members are the bitterness gene in the bottle gourd. Of the 89 predicted genes in the *QBt.2* region, there is a transcription factor bHLH35-like isoform X1 gene and three cytochrome P450 CYP73A100-like genes, which might be functionally related to known cucurbit bitterness genes. More future work is required to draw a clearer picture on the causal genes underlying bitterness QTLs in the bottle gourd.

Our results shed new light on the molecular genetic mechanisms underlying fruit bitterness in the bottle gourd. From the perspective of breeding, it will be useful to guide the selection of parental lines to avoid the occurrence of bitter fruits in breeding programs. Breeders could use the flanking markers to *QBt.1* and *QBt.2* to screen their germplasm and breeding lines to predict if the hybrids would show a bitterness phenotype without the need of making real crosses.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

XYW, PX and GL conceived and designed the research. XYW performed the experiments and wrote the manuscript. XHW constructed the population and collected bitterness phenotypes. YW, BW and ZL carried out the field work. All authors analyzed the data and read and approved the final manuscript. PX revised the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01493/full#supplementary-material

SUPPLEMENTARY TABLE 1 | Genotype data and phenotype data for QTL mapping in the F$_2$ population.

SUPPLEMENTARY TABLE 2 | Colinearity analysis of the cucurbitacin biosynthetic and regulatory genes in four cucurbits.

REFERENCES

Andeweg, J. M., and DeBruyn, J. W. (1959). Breeding of non-bitter cucumbers. *Euphytica* 8, 13–20. doi: 10.1007/BF00022084

Balkema-Boomstra, A. G., Zijlstra, S., and Verstappen, F. W. (2003). Role of cucurbitacin C in resistance to spider mite (*Tetranychus urticae*) in cucumber (*Cucumis sativus* L.). *J. Chem. Ecol.* 29, 225–235. doi: 10.1023/A:1021945101308

Beevy, S. S., and Kuriachan, P. (1996). Chromosome numbers of south Indian Cucurbitaceae and a note on the cytological evolution in the family. *J. Cytol. Genet.* 31, 65–71.

Chen, J. C., Chiu, M. H., Nie, R. L., Cordell, G. A., and Qiu, S. X. (2005). Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Nat. Prod. Rep.* 22, 386–399. doi: 10.1039/b418841c

Crawford, G. W. (1992). *Origins of Agriculture: An International Perspective* Vol. pp. Cowan, CW, and Watson, PJ editors. Washington, DC: Smithsonian Institution Press, 7–38.

Davis, A. R., Perkins-Veazie, P., Sakata, Y., López-Galarza, S., Maroto, J. V., Lee, S., et al. (2008) Cucurbit grafting. *CRC. Crit. Rev. Plant Sci.* 27, 50–74. doi: 10.1080/07352680802053940

Erickson, L. D., Smith, D. B., Clarke, C. A., Sandweiss, H. D., and Tuross, N. (2005). An Asian origin for a 10, 000-year-old domesticated plant in the Americas. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18315–18320. doi: 10.1073/pnas.0509279102

King, S. R., Davis, A. R., Liu, W., Levi, A., King, S. R., Davis, A. R., et al. (2008). Grafting for disease resistance. *HortScience* 43, 1673–1676. doi: 10.21273/HORTSCI.43.6.1673

Lester, G. (1997). Melon (*Cucumis melo* L.) fruit nutritional quality and health functionality. *HortTechnology* 7, 222–227. doi: 10.21273/HORTTECH.7.3.222

Matsuo, K., DeMilo, A. B., Schroder, R. F. W., and Martin, P. A. W. (1999). Rapid high-performance liquid chromatography method to quantitate elaterinide in juice and reconstituted residues from a bitter mutant of Hawkesbury watermelon. *J. Agric. Food Chem.* 47, 2755–2759. doi: 10.1021/jf9811572

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., and Chen, H. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Van Ooijen. (2004). *MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations*. Wageningen, Netherlands: Kyazma B.V.

Wang, Y., Xu, P., Wu, X., Wu, X., Wang, B., and Huang, Y. (2018). GourdBase: a genome-centered multi-omics database for the bottle gourd (Lagenaria siceraria), an economically important cucurbit crop. *Sci. Rep.* 8, 3604. doi: 10.1038/s41598-018-22007-3

Wu, W., and Huang, B. (2006). The strategy for mapping interaction genes using the bulked segregant analysis and Mapmaker/Exp software. *Chin. Sci. Bull.* 51 (18), 2134–2138. doi: 10.1007/s11434-006-2174-1

Wu, X., Xu, P., Wu, X., Wang, B., Lu, Z., and Li, G. (2017a). Genome-wide association analysis of free glutamate content, a key factor conferring umami taste in the bottle gourd [*Lagenaria siceraria* (Mol.) Standl.]. *Sci. Hortic.* 225, 795–801. doi: 10.1016/j.scienta.2017.08.015

Wu, X., Xu, P., Wu, X., Wang, B., Lu, Z., and Li, G. (2017b). Development of insertion and deletion markers for bottle gourd based on restriction site-associated DNA sequencing data. *Hortic. Plant J.* 3 (1), 13–16. doi: 10.1016/j.hpj.2017.01.011

Wu, S., Shamimuzzaman, M., Sun, H., Salse, J., Sui, X., and Wilder, A. (2017c). The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a *Papaya* ring-spot virus resistance locus. *Plant J.* 92, 963–975. doi: 10.1111/tpj.13722

Xu, P., Wu, X., Luo, J., Wang, B., Liu, Y., Ehlers, J. D., et al. (2011). Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. *BMC Genomics* 12, 467. doi: 10.1186/1471-2164-12-467

Xu, P., Xu, S., Wu, X., Tao, Y., Wang, B., Wang, S., et al. (2014). Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *Plant J.* 77, 430–442. doi: 10.1111/tpj.12370

Zhang, S., Miao, H., Sun, R., Wang, X., Huang, S., Wehner, T. C., et al. (2013). Localization of a new gene for bitterness in cucumber. *J. Hered.* 104 (1), 134–139. doi: 10.1093/jhered/ess075

Zhang, G. (1981). Interaction of genes and the expression of bitterness in *Lagenaria Sicereria. Acta Hortic. Sin.* 8, 43–48.

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., and Wang, H. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* 2, 16183. doi: 10.1038/nplants.2016.183

Genetic Mapping and Discovery of the Candidate Gene for Black Seed Coat Color in Watermelon (*Citrullus lanatus*)

Bingbing Li[†], Xuqiang Lu[†], Haileslassie Gebremeskel, Shengjie Zhao, Nan He, Pingli Yuan, Chengsheng Gong, Umer Mohammed and Wenge Liu[*]

Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, Henan, China

Seed coat color is an important trait highly affecting the seed quality and flesh appearance of watermelon (*Citrullus lanatus*). However, the molecular regulation mechanism of seed coat color in watermelon is still unclear. In the present study, genetic analysis was performed by evaluating F_1, F_2 and BC_1 populations derived from two parental lines (9904 with light yellow seeds and Handel with black seeds), suggesting that a single dominant gene controls the black seed coat. The initial mapping result revealed a region of interest spanning 370 kb on chromosome 3. Genetic mapping with CAPS and SNP markers narrowed down the candidate region to 70.2 kb. Sequence alignment of the three putative genes in the candidate region suggested that there was a single-nucleotide insertion in the coding region of *Cla019481* in 9904, resulting in a frameshift mutation and premature stop codon. The results indicated that *Cla019481* named *ClCS1* was the candidate gene for black seed coat color in watermelon. In addition, gene annotation revealed that *Cla019481* encoded a polyphenol oxidase (PPO), which involved in the oxidation step of the melanin biosynthesis. This research finding will facilitate maker-assisted selection in watermelon and provide evidence for the study of black seed coat coloration in plants.

Keywords: watermelon, seed coat color, polyphenol oxidase, genetic mapping, candidate gene

INTRODUCTION

Watermelon [*Citrullus lantus* (Thunb.) Matsum. & Nakai (2n = 2x = 22)] is an important horticultural crop worldwide (Mohr, 1986; Levi et al., 2001). Seed coat color is an essential part of botanical character in watermelon which plays an important role in watermelon breeding, especially for seeded watermelon (Mavi, 2010). However, there are few molecular studies for seed coat color compared to other traits (Yan et al., 1996; Łopusiewicz, 2018). Seed coat color is also associated with the biochemical characteristics of seeds, the amounts and activities of antioxidants (El-Bramawy et al., 2008), and affects the flesh appearance (Wehner, 2008). Black seed is attractive matched with red canary yellow flesh, while white or light seed color is an ideal character for the near-seedless cultivars (Wehner, 2008).

Seed coat has a large diversity of colors in different species, genotypes and even at different seasons or developmental stages (Demir et al., 2004; Wan et al., 2016). The mechanism of seed coat

coloration has been well-illuminated in Brassicaceae family, especially in the model species of *Arabidopsis* and its close phylogenetic relative *Brassica* (Debeaujon et al., 2000; Baxter et al., 2005; Li et al., 2010; Yu, 2013; Padmaja et al., 2014). Seed coat color in *Arabidopsis* and *Brassica* is generally categorized into two main classes, yellow and brown (Yu, 2013). The yellow seed is formed due to the transparent and colorless seed coat, resulting in the exposure of yellow embryo (Sagasser et al., 2002; Bharti and Khurana, 2003). Flavonoids (flavonols and proanthocyanidin) are the main components for the seed coat coloration in *Arabidopsis* and *Brassica*. The endothelium layers synthesize proanthocyanidins (PAs) that condense into tannins and oxidize, resulting in a brown color in mature seeds (Yu, 2013).

In recent years, dozens of genes associated with flavonoid biosynthesis have been well characterized by TRANSPARENT TESTA (*tt*) mutants (Debeaujon et al., 2000; Baxter et al., 2005; Li et al., 2010; Yu, 2013; Padmaja et al., 2014). Nearly 27 mutations have been detected to show association with the flavonoid pathway related to the seed coat color in *Arabidopsis* (Yu, 2013). Several genes identified at the molecular level have been categorized into two groups of structural proteins: the early biosynthetic genes (EBGs) including chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavanone-3-hydroxylase (*F3H*), and flavanone-3'-hydroxylase (*F3'H*); the late biosynthetic genes (LBGs) including dihydroflavonol reductase (*DFR*), leucocyanidin dioxygenase (*LDOX*) and anthocyanidin reductase (*ANR*). The expression of the underlying biosynthetic genes was regulated by some regulatory factors (*TT1, TT2, TT8, TT16, TTG1, TTG2, PAP1, GL3, ANL2, FUSCA3, KAN4*) (Qu et al., 2013; Yu, 2013; Padmaja et al., 2014). In addition to flavonoid, it was reported that melanin also affected the seed coat color of seeded rapes (Zhang et al., 2006). The seed coat color mainly depends on the content of melanin in the last stage of seed development (Zhang et al., 2006; Yu, 2013). Melanin is a widespread black compound in nature, especially in the seed coat (Park and Hoshino, 2012; Wan et al., 2016). Natural melanin has potential values for pharmacology, cosmetics and functional foods due to its antioxidant, radio-protective, antitumor, antiviral, antimicrobial and anti-inflammatory properties (Łopusiewicz, 2018). However, there are limited studies about the candidate genes regulating melanin biosynthesis controlling the seed coat coloration, and the genetic regulatory mechanisms accounting for black seed coat coloration has not been well elucidated (Yu, 2013).

Polyphenol oxidase (PPO) is the key enzyme catalyzing phenolic compounds to form o-quinones, which could easily react with amines, proteins or other phenols to produce dark melanin pigments (Mayer and Harel, 1979; Matheis and Whitaker, 1984; Whitaker and Lee, 1995; Yu, 2013). In higher plants, it has been reported that PPO is responsible for the dark color of damaged kernels, fruits, or vegetables, which may be involved in disease resistance (Yu et al., 2008). *SiPPO* encoding polyphenol oxidase was proposed to be responsible for generating black or whiter sesame (Wei et al., 2015).

In watermelon, previous studies for seed coat color mainly focused on the inheritance, antioxidant of the seed coat color and the extraction of pigment. Watermelon exhibits a wide range of seed coat colors, commonly white, tan, brown, black, red, green, and dotted (Poole, 1941). McKay (1936) suggested that three genes determined watermelon seed coat color: *r*, *w*, and *t* for red, white, and tan, respectively. The interaction of the three genes produced six patterns: black (*RR TT WW*), clump (*RR TT ww*), tan (*RR tt WW*), white with tan tip (*RR tt ww*), red (*rr tt WW*), and white with pink tip (*rr tt ww*) (Kanda, 1931; McKay, 1936; Poole, 1941). A modifier, *d* was suggested to produce a black and dotted seed coat (Poole, 1941). Paudel et al. (2019) developed three segregating F_2 populations to re-investigate the four-gene model and map the locus of the four genes. Paudel et al. (2019) found that the inheritance of the *T* locus did not fit the four-gene model in F_2 progenies (dotted black × red). The tannish seed coat color was observed and classified as tan[1] that was affected by T^1, a novel locus or a different allele of the *T* locus (Poole, 1941; Paudel et al., 2019). Besides, the *R*, *T*, *W*, and *D* loci were mapped on chromosomes 3, 5, 6 and 8, respectively. In watermelon, melanin accumulation in seed coat was responsible for the black seed coloration and black seeds were considered as a potential source of natural melanin (Yan et al., 1996; Łopusiewicz, 2018).

Until now, there are no concrete molecular evidences in watermelon seed coat coloration. Studies on the melanin pigmentation in plant seed coloration are not plentiful. In this study, we aimed to illustrate the inheritance of watermelon seed coat color and detect the candidate gene responsible for black seed coat coloration. The present study will facilitate marker-assisted selection and provide additional evidence for the melanin pigmentation in plant seed coat.

MATERIALS AND METHODS

Plant Materials and Genetic Mapping Population

The preliminary mapping population consisted of 126 recombinant inbred lines (RILs, F_7), derived from a cross between the inbred lines 9904 (female parent) and Handel (male parent) which have light yellow and black seed coat, respectively (Li et al., 2018). In the segregating population, there was a third phenotype of seed coat color named dotted, which was not completely black (black strip or dot) seed coat (Poole, 1941; Li et al., 2018). The F_2 population was used to perform fine mapping. The backcross population was produced by hybridizing F_1 plant with each parent to create BC_1P_1 (F_1×9904) and BC_1P_2 (F_1×Handel) and used to validate genetic inheritance of seed coat color.

For the genetic map construction and segregation analysis, the RIL population was grown together with the parental lines at two locations under three environments: Sanya Experimental Station in 2016 (Hainan, open field) and Xinxiang Experimental Station in 2017 (Henan, greenhouse and open field) (Li et al.,

2018). The F_2 population was grown in 2018 (Henan) during spring season with 560 individuals. The 140 BC_1P_1 and 161 BC_1P_2 individuals were grown in 2018 autumn season (**Table S1**). The phenotype of seed coat color was determined by visual observation, and watermelon seeds were categorized into black, light yellow and dotted groups based on their appearance at 40 days after pollination (DAP).

Measurement of Melanin, Polyphenol and Flavonoid Content

The mature seed coat of the parental lines at 40 DAP were collected and pigments (melanin, polyphenol and flavonoid) were extracted according to the methods described by Ye et al. (2001a) and Li et al. (2011). Three biological and technical replicates were used for the measurement. Average content for each sample was calculated.

Determination of Polyphenol Oxidase Activity

Seed coat samples (18 and 26 DAP) without the cotyledon and embryo were collected from parental lines by scalpels, and then were used to measure the activity of PPO. PPO was extracted and measured using the PPO Assay Kit (Solarbio, China) according to the manufacturer's instruction.

DNA Extraction, Analysis of Whole-Genome Resequencing Data, and Genetic Map Construction

Young leaves from two parental lines and the segregating population were collected and stored at −80°C until DNA was extracted. The genomic DNA was extracted by the cetyltrimethyl ammonium bromide (CTAB) method (Murray and Thompson, 1980). DNA was quantified with a NanDrop-1000 spectrophotometer (NanoDrop, USA) and was evaluated by electrophoresis in 1.0% agarose gel.

In our previous study (Li et al., 2018), we constructed a high-density genetic map based on whole-genome resequencing of the RIL population and both parental lines. A total of 7.67 Gbp, 8.81 Gbp, and 177.08 Gbp of high-quality reads were obtained from the 9904, Handel, and RIL population, respectively. The average coverage depths of the markers were 19-fold for the male parent, 17-fold for the female parent, and 3-fold for the RIL population. The average Q30 ratio was more than 85%, and the average GC content was nearly 35% for the RIL individuals (Li et al., 2018). The distribution of SNP mutations and coverage of assembly scaffolds by high-quality reads indicated that the genome resequencing was sufficiently random (Li et al., 2018). A total of 178,762 SNPs with at least a 4-fold sequencing depth were obtained by analyzing the parental lines. All of the SNP sites in the RILs were integrated into a recombination bin unit, and 2,132 recombinant bins comprising 103,029 SNPs were used to construct the genetic map (Li et al., 2018). As compared to other recent studies in watermelon, we found that greater number of SNP markers were mapped to this genetic map (Li et al., 2018). The final high-density genetic map had a total length of

1,508.94 cM, with an average distance of 0.74 cM between adjacent bin markers. Additionally, the haplotype, heat maps and collinearity of the genetic map with watermelon reference genome showed that the high-density genetic map was accurately assembled with good quality (Li et al., 2018). The LOD thresholds for determining significant loci were estimated from 1,000 permutations and a minimum LOD score of 2.5 was used to judge the presence of loci on the chromosome (Churchill and Doerge, 1994).

Molecular Marker Development and Genetic Mapping

Re-sequenced data were compared with the available '97103' watermelon reference genome version 1 from the Cucurbit Genomics Database (http://cucurbitgenomics.org/) to identify reliable SNPs through a filter pipeline (Takagi et al., 2013). To narrow down the candidate region and verify the accuracy of the preliminary mapping derived from the genetic map, the corresponding cleaved amplified polymorphic sequence (CAPS) markers were developed based on SNPs (**Table S2**). Finally, 57 CAPS makers were developed to screen F_2 population (540 individuals) for fine mapping (**Table S2**).

PCR amplification was performed in a 10 μl reaction with 1 μl DNA, 5 μl PCR master mix, 0.5 μl of 10 μM per primer, and 3 μl distilled water. The PCR protocol was performed under the following conditions: initial denaturation at 94°C for 1 min and 30 s; followed by 30 cycles at 94°C for 20 s, 57°C for 20 s, 72°C for 50 s; and a final extension at 72°C for 5 min. Then, the corresponding restriction endonucleases were used to digest the amplified PCR products at 37°C or 65°C for 4–10 h following the manufacturer's instructions. The digested products were separated on 1.0% agarose gels and visualized with a Versa Doc (Bio-Rad). The markers with polymorphisms were used for fine mapping.

RNA Isolation and Quantitative Real-Time PCR Analysis of the Candidate Gene

The seed coat samples from different developmental stages (18 and 26 DAP) and other tissue samples, including roots, stems, leaves, and male flowers were collected from both parental lines. RNA was isolated using the plant total RNA purification kit (TIANGEN, China) according to the manufacturer's instructions and then the first-strand cDNA was synthesized using a cDNA synthesis kit (Takara, Japan).

The gene-specific primers of the candidate genes and reference gene *Actin* (Kong et al., 2015) for quantitative real-time PCR (qRT-PCR) were designed based on the Cucurbit Genomic Database (http://cucurbitgenomics.org), using the software Primer Premier 5. The expression levels of the candidate genes were performed using a LightCycler480 RT-PCR system (Roche, Swiss) with a Real Master Mix (SYBR Green) kit (Toyobo, Japan). Amplification was carried out in a 20 μl reaction mixture containing 1 μl cDNA, 1 μl forward and reverse primers (10 μM), 10 μl 2 × SYBR Green real-time PCR mixes, with nuclease-free water added to a total reaction of 20 μl.

Three biological and technical replicates were used for qRT-PCR. Average relative expression levels for each sample were calculated. The expression level was analyzed by the $2^{-\triangle\triangle Ct}$ method (Livak and Schmittgen, 2001), and the primer sequences used in this study are listed in **Table S3**.

Sequence and Phylogenetic Analysis of the Candidate Gene

The sequence and gene function were retrieved from the Cucurbit Genomics Database (http://cucurbitgenomics.org). DNA and amino acid sequences were aligned using DNAMAN (version 9). Phylogenetic analysis were performed using MEGA 7 software with a bootstrap method and 1000 replications (Kumar et al., 2016).

RESULTS

Inheritance and Phenotypic Characterization of Seed Coat Color in Watermelon

In the present study, the parental lines showed significant variations in seed coat color. 9904 has light yellow seed coat, while Handel has black seed coat color. In F_1 population, all the seeds were black without segregation, which revealed that black was dominant to light yellow seed coat color. There was a third phenotype of seed coat color named dotted, which was not completely black (black strip or dotted) seed coat in the F_2 population (Poole, 1941; Li et al., 2018). The F_2 population separated into 332 plants with black seed coat, 101 plants with

dotted seed coat and 127 plants with light yellow seed coat, resulting in good fit to a 9:3:4 segregation ratio ($\chi^2 = 2.28$, P = 0.32) (**Table S1**). For the BC_1P_1 population, there were 40 plants with black seed coat, 32 plants with dotted seed coat and 68 plants with light yellow seed coat, showing a ratio of 1:1:2 ($\chi^2 = 1.03$, P = 0.60). While all the 161 individuals of BC_1P_2 had black seed coat. When seeds from the F_2 and BC_1 generation were scored as black (black and dotted) vs nonblack (light yellow), the seed phenotype fits a typical Mendelian segregation ratio of 3:1 ($\chi^2 = 1.61$, P = 0.2) and 1:1 ($\chi^2 = 0.11$, P = 0.73), respectively. According to the genetic analysis, we detected that two genes account for the seed coat color in present materials resulting in three different phenotypes, black (*W D*), dotted (*W d d*), and light yellow (*w w*). The *W* gene was responsible for the black coloration, while *d* accounted for the distribution of the black in seed coat resulting in partial black named dotted. Besides, when seeds were scored as black (black and dotted) and nonblack (light yellow), the phenotype of the seed coat perfectly fit a ratio of Mendelian single gene segregation ratio, black (*W*): light yellow (*w w*), which indicated that black seed coat is controlled by a single dominant gene (*W*).

Genetic Mapping of the Candidate Gene

In our previous study, we detected a prominent locus (qsc-c3-1) associated with black seed coat color (Li et al., 2018). To narrow down the genetic region and identify the candidate genes for watermelon seed coat color, the ampliative F_2 population with 560 individuals was developed. Based on the '97103' watermelon reference genome version 1 (http://cucurbitgenomics.org), 57 CAPS and 10 SNP markers were developed in the candidate region on chromosome 3 to screen all F_2 individuals for
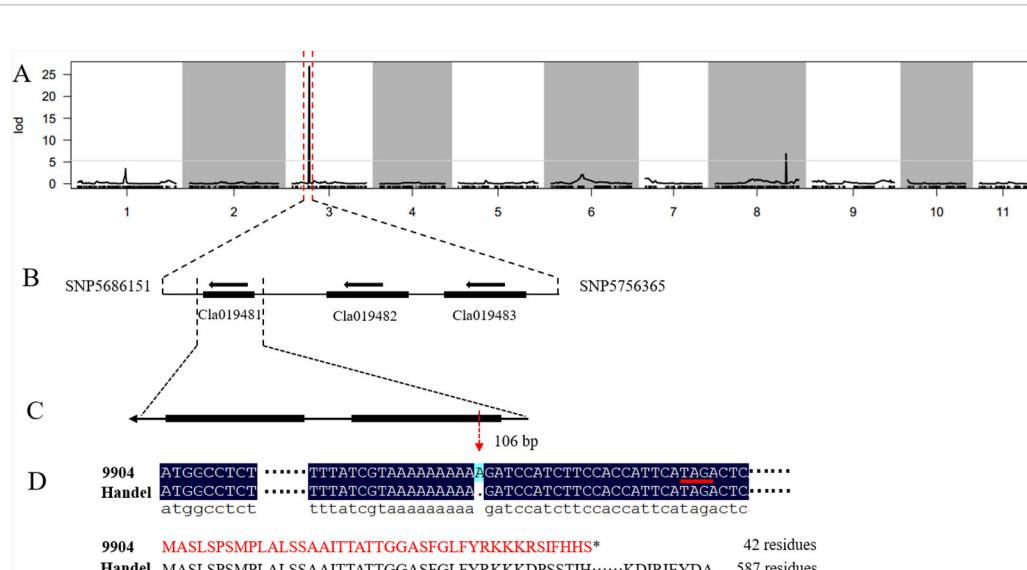
FIGURE 1 | Genetic mapping of the seed coat color gene and candidate gene analysis in watermelon. **(A)** The candidate gene was located on chromosome 3. **(B)** The candidate gene was localized to a 70.2 kb region between the flanking markers SNP5686151 and SNP5756365. **(C)** Structure of the *Cla019481* gene. **(D)** A frameshift mutation and early termination of translation in 9904.

TABLE 1 | Phenotypes and genotypes of recombinant individuals showing the recombinant breaking points.

No.	Phenotype	Chr3_5654223	SNP5666004	SNP5686151	SNP5707786	SNP5733877	SNP5756365	SNP5822743	SNP5832388	SNP5848889
1	LY	a	a	a	a	a	a	a	a	a
2	LY	a	a	a	a	a	a	a	a	a
3	LY	H	a	a	a	a	a	a	a	a
4	B	H	H	H	H	H	H	H	H	H
5	B	a	a	a	A	H	H	H	H	H
6	LY	H	a	a	a	a	a	a	a	a
7	B	a	a	a	A	H	H	H	H	H
8	B	H	H	H	H	H	a	a	a	a
9	B	a	A	A	A	A	A	A	A	A
10	B	a	A	A	A	A	A	A	A	A
11	B	A	A	A	A	A	A	a	a	a
12	B	A	A	A	A	H	a	a	a	a
13	LY	A	A	A	a	a	a	a	a	a

LY indicates light yellow seed coat color. B indicates the black (black and dotted) seed coat color. The alleles are abbreviated according to their origin: a, 9904 (Light yellow); A, Handel (Black); H, heterozygous. The highlighted red means the final mapping interval.

polymorphic analysis (**Table S2**). Finally, the candidate gene was delimited to a 70.2 kb region between SNP5686151 and SNP5756365 with five recombinant individuals (**Figure 1**, **Table 1**). The candidate gene responsible for black seed coat color was delimited in to a nearly 70.2 kb interval on chromosome 3 (**Figure 1**).

Determination of Pigment Contents and PPO Activity

The melanin, polyphenol and flavonoid content of seed coat samples from the two parental lines at 40 DAP were measured. The results showed that black seed coat contained significantly higher (~4.4 folds) melanin content compared with light yellow seed coat (**Figure 2**). While light yellow seed coat contained higher polyphenol (~5.6 folds) and flavonoid (~8.1 folds) contents than black seed coat (**Figure 2**). PPO activity in black seed coat was 2.5 and 2.7-folds higher as compared to light yellow seed coat at 18 DAP and 26 DAP, respectively (**Figure 2**).

Sequence and Annotation Analysis of the Candidate Genes

Based on the narrowed interval and '97103' watermelon reference genome version 1 (http://cucurbitgenomics.org/), only three putative genes (*Cla019481*, *Cla019482*, *Cla019483*)
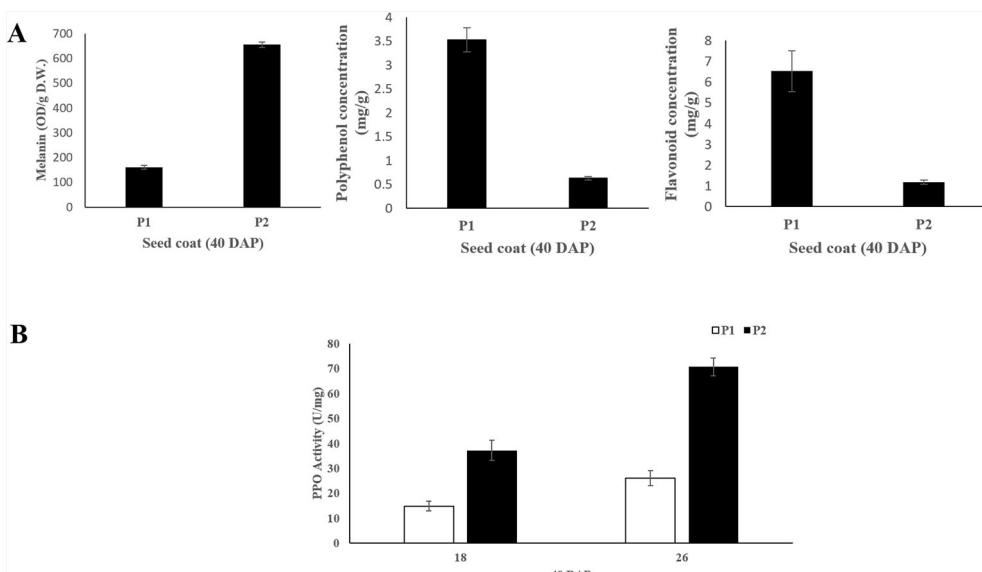
FIGURE 2 | (A) Melanin, polyphenol and flavonoid content of seed coat at 40 DAP between the two parental lines. **(B)** PPO activity of the seed coat at 18 and 26 DAP between the two parental lines. Three biological and technical replicates were used for the measurement.

were found in the 70.2 kb region (**Figure 1**). To analyze the gene sequence of the candidate genes and detect the gene responsible for seed coat color in watermelon, we designed gene specific primers to clone the whole gene and the entire coding sequence (CDS) from both parental lines (**Table S3**). The sequence alignment of the three genes between 9904 and Handel showed a base insertion existed in the CDS region of *Cla019481* at the position 106 bp in 9904 (light yellow) resulting in a frameshift mutation and leading to early termination of translation (42 residues) (**Figure 1**). A synonymous SNP mutation in the exon region of *Cla019482* leading to no amino acid change was detected. Besides, there was no sequence variation found in *Cla019483* between the two parental lines. Moreover, we have sequenced the 5'-upstream sequence from CDS region for about 2,000 bp to analyze the promoter region of the three candidate genes. There were no variations for *Cla019482* and *Cla019483* between the two parental lines, and SNPs were detected in *Cla019481* between 9904 and Handel. To verify the insertion mutation, 100 individuals were selected from the F_2 population to check the consistency of the phenotypes and genotypes. The results indicated that 74 individuals with black seed coat (black and dotted) were homozygous dominant or heterozygous, and 26 individuals with nonblack seed coat were homozygous recessive. The phenotypes were perfectly consistent with the genotypes. Therefore, we proposed that *Cla019481* named *ClSC1* was the candidate gene for black seed coat color in watermelon.

The Cucurbit Genome Database (http://cucurbitgenomics. org/) and BLAST of the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) were used to predict the function of the candidate genes. The BLAST results indicated that all three genes in the candidate region were predicted to encode polyphenol oxidase proteins. PPO is the key enzyme to catalyze the phenolic compounds to o-quinones, which can polymerize to form the black melanin pigments attributed to black coloration during seed development (Mayer and Harel, 1979; Matheis and Whitaker, 1984; Whitaker and Lee, 1995; Yu, 2013). According to this result, *Cla019481* is considered to be the candidate gene accounting for the black seed coat in watermelon.

Candidate Gene Expression Analysis

Seed coat samples at different development stages (18 and 26 DAP) and other tissues including roots, stems, leaves and male flowers were collected to perform expression analysis of *Cla019481* using qRT-PCR. We observed that *Cla019481* was only significantly expressed in seed coat at Handel (black seed coat; **Figure 3**). These results indicate that the different expression levels of *Cla019481* in the seed coat between the two parental lines might result in the different seed coat coloration in watermelon.

Phylogenic and Protein Domain Analysis

To better understand the relationship between *Cla019481* protein and its homologs, we used BLAST search in NCBI
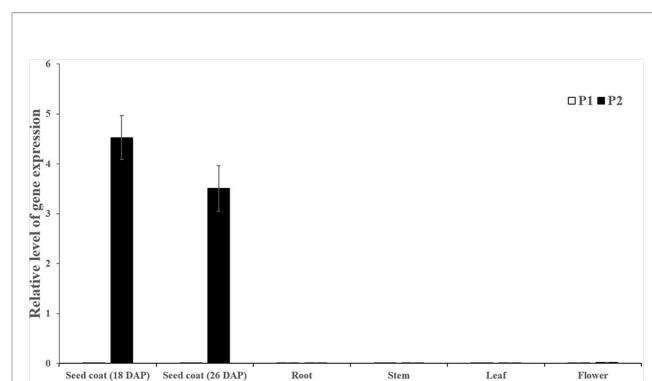
FIGURE 3 | Quantitative real-time PCR analysis of *Cla019481* in different development stages (18 and 26 DAP) and different tissues. Three biological and technical replicates were used for qRT-PCR.

database (https://www.ncbi.nlm.nih.gov/) and MEGA 7 software to perform phylogenic analysis using an alignment of the closest homologs through bootstrap method with 1,000 replications (Kumar et al., 2016). The resulting neighbor-joining tree showed that *Cla019481* from watermelon had the closest phylogenetic relationship to the homolog from *Momordica charantia* and grouped together with homologs from Cucurbitaceae family, including *Cucurbita moschata*, *Cucurbita pepo subsp. pepo*, *Luffa aegyptiaca*, *Cucumis melo*, and *Cucumis sativus* (**Figure S1**), which revealed that *Cla019481* was evolutionarily conserved within the Cucurbitaceae family.

The *Cla019481* protein domain was generated with the online Pfam database (http://pfam.xfam.org). The result showed that there were three unique domain architectures, belonging to tyrosinase, PPO1 DWL and PPO1 KFDV family, respectively (**Figure S1**). The single base insertion in CDS of 9904 resulted in a premature protein translation encoding a truncated polypeptide (42 amino acids), which is likely to be non-functional. The amino acid alignment analysis *via* NCBI and Rice Genome (http://rice.plantbiology.msu.edu/index.shtml) revealed that the *ClSC1* protein shared 44.36% sequence identity with *Phr1* (*LOC_Os04g53300*) in *Oryza sativa*, and both shared the same tyrosinase, PPO1 DWL and PPO1 KFDV domains (Yu et al., 2008).

DISCUSSION

The development of the high-throughput sequencing technology and the assembly of watermelon reference genome provided a powerful tool to identify loci associated with important traits (Guo et al., 2013; Wan et al., 2016). However, the marker density was far from saturated for marker-assisted selection (MAS) or for cloning important genes (Ren et al., 2015). To increase marker saturation and develop marker resources for watermelon, we constructed a high-density genetic map based on whole genome resequencing of the RIL

population derived from a cross between 9904 and Handel (Li et al., 2018). We detected a prominent locus for black seed coat color in an approximate 370 kb region on chromosome 3. In the present study, we performed genetic mapping and identified a candidate gene *Cla019481* named as *ClSC1* accounting for black seed coat coloration in watermelon.

Watermelon seed coat color ranged from almost pure white to red, green, brown, tan, mahogany, and black in various superimposed complex patterns (Poole, 1941). The genetic pattern of seed coat color in watermelon is intricate and attractive. Seed coat color in watermelon was controlled by multi-genes where a certain color was dominant over others (Maina and Nyambura, 2018). According to Kanda (1931), there were at least seven genes controlling the phenotypes of seed coat color with black dominant to all colors. A three-gene model has been proposed to control the seed coat color: *r* (red), *w* (white), and *t* (tan), respectively (McKay, 1936; Poole, 1941). The interaction of these genes produced six base colors: black *(RR TT WW)*; clump *(RR TT ww)*; tan *(RR tt WW)*; white with tan tip *(RR tt ww)*; red *(rr tt WW)*; and white with pink tip *(rr tt ww)* (Kanda, 1931; McKay, 1936; Poole, 1941). Furthermore, a modified gene, *d* was proposed to be responsible for a black dotted seed coat (Poole, 1941; Hawkins et al., 2001). A single pair of gene, c_r was reported to account for the formation of the cracks on the seed coat (Abd el-Hafez et al., 1985). Paudel et al. (2019) found that T^1 locus was a different allele or novel locus than the previously described *T* locus and developed markers UGA3_5820134, UGA5_4591722, UGA6_7076766, and UGA8_22729513 for MAS of seed coat color in watermelon.

The phenotype of watermelon seed coat color is difficult to be classified because of the restriction in materials and various degrees of segregation. Since seed coat color is under multigenic control, advanced generations, such as recombinant inbred lines and reciprocal backcrosses, are needed to illuminate its complex inheritance and molecular mechanisms (Abdel-Hafez et al., 1985). We constructed a RIL population from a cross between inbred lines, 9904 with light yellow seeds and Handel with light yellow seeds. As compared to the other materials for seed coat color, the genetic background is relatively pure and there are only three phenotypes in the progeny (Abd el-Hafez et al., 1985; Li et al., 2018). According to the genetic analysis and previous literatures, we detected that two genes account for the seed coat color in present materials resulting in three different phenotypes, black (*W D*), dotted (*W d d*), light yellow (*w w*). The *W* allele was responsible for the black coloration, while *d* accounted for the distribution of the black in seed coat resulting in partial black named dotted. Besides, when seeds were scored as black (black and dotted) and nonblack (light yellow), the phenotype of the seed coat perfectly fit with Mendelian single gene segregation ratio, black (*W*): light yellow (*w w*), which indicated that black seed coat is controlled by a single dominant gene (*W*).

Seed coat color in watermelon is not only an important commercial trait affecting the quality especially for seeded watermelon, but also associated with water uptake, seed

dormancy and germination (Mavi, 2010). Until now, few studies have been reported on QTLs or candidate genes responsible for seed coat color in watermelon and the regulatory mechanisms of seed coat coloration remains elusive. In this study, we developed a dilated F_2 population and performed genetic mapping by CAPS and SNP molecular markers. Finally, the causal gene was delimited to a 70.2 kb region between SNP5686151 and SNP5756365 on chromosome 3. According to the watermelon reference genome, only three putative genes were annotated in this interval. The sequence alignment between the parental lines showed that one single-nucleotide base insertion existed in the CDS region of *Cla019481* at the position 106 bp in 9904 (light yellow). The SNP mutation resulted in a frameshift mutation and led to early termination of translation (42 amino acids), that likely to led to a non-functional protein (**Figure 1**). This result indicated that *Cla019481* named as *ClSC1* was the candidate gene for watermelon seed coat.

The formation and accumulation of pigments, including polyphenols, anthocyanin, flavonoid, and melanin, affect the color of seed coat (Ye et al., 2001b; Ye et al., 2001c; Zhang et al., 2006). The mechanism of seed coat coloration was well-studied in *Arabidopsis* and *Brassica* species using deficient mutants (Yu, 2013). Flavonoid (flavonols and proanthocyanidin) was responsible for the pigmentation pattern of seeds in *Arabidopsis* and *Brassica*. Dozens of genes associated with the flavonoid biosynthetic pathway have been detected using *Arabidopsis* transparent testa mutations (Yu, 2013; Appelhagen et al., 2014). Twenty-three genes have been detected at the molecular level, including enzymes (*CHS*, *CHI*, *F3H*, *F3'H*, *DFR*, *LDOX*, *FLS*, *ANR*, *LACCASE*), transports (*TT12*, *TT19*, *AHA10*), and regulatory factors (*TT1*, *TT2*, *TT8*, *TT16*, *TTG1*, *TTG2*, *PAP1*, *GL3*, *ANL2*, *FUSCA3*, *KAN4*) (Debeaujon et al., 2000; Baxter et al., 2005; Li et al., 2010; Yu, 2013; Padmaja et al., 2014). Due to the close phylogenetic relationship between *Arabidopsis* and *Brassica*, a number of *Brassica tt* orthologs in *Arabidopsis* have been cloned (Yu, 2013). However, compared to *Arabidopsis,* the *tt* mutants in seed coat coloration for *Brassica* do not show significant effects on morphologic and physiological performance, suggesting the presence of function complement for duplicated genes or side-pathway. There were some *Brassica* with dark brown or black seed coats indicating that the coloration in higher plants might be more complex (Yu, 2013).

In addition to flavonoids, melanin was also an important pigment for seed coat coloration, but the mechanism for seed coat coloration is still elusive and the literatures are not plentiful (Marles and Gruber, 2004; Yu, 2013). Zhang et al. (2006) reported that polyphenols, anthocyanin and flavonoid were mainly responsible for coloration in the early and middle developmental stages of the black and yellow rape-seed, but the color was mainly affected by melanin in the late stage. PPO associated with the conversion of phenolic compounds was the key enzyme in the melanin pathway (Mayer and Harel, 1979). PPO catalyzes two steps of enzymatic reactions. The first leads the ortho-hydroxylation of monophenols to ortho-diphenols.

The second procedure is the oxidation of ortho-diphenols to ortho-quinone, which easily undergoes non-enzymatic reactions to form dark melanin pigments (Mayer and Harel, 1979; Matheis and Whitaker, 1984; Whitaker and Lee, 1995). PPO is the major enzyme resulting in the browned and darkened fruits, vegetables and cereals grains (Yan et al., 2017). However, the correlation between PPO and black seed coat coloration has not been well characterize. It was reported that there was significant positive correlation between PPO activity and melanin and significant negative correlation with polyphenol in *Brassica napus* seed coat (Ye et al., 2001a; Ye et al., 2002). Wei et al. (2015) proposed that *SiPPO* encoding polyphenol oxidase was responsible for generating black or whiter sesame.

PPO genes have been cloned and illustrated in different plant species, including tomato (Shahar et al., 1992), potato (Hunt et al., 1993), grape (Dry and Robinson, 1994), apple (Boss et al., 1995), wheat (Demeke and Morris, 2002), rice (Yu et al., 2008), barley (Taketa et al., 2010), broad bean (Taketa et al., 2010). However, no PPO-like genes were identified in *Arabidopsis* and chlorophyte (green algae) (Tran et al., 2012). All of the PPO genes have two conserved copper-binding domains (CuA and CuB), forming the central domain of the catalytic site (He et al., 2007). In rice, *Phr1* encoding PPO contributed the discoloration of hulls and coarse grains of *indica*-type cultivars, while the nonfunctional *Phr1* leaded to no discoloration of *japonica* grains (Yu et al., 2008). The insertion or deletion resulting in frameshift mutations in the *Phr1* CDS accounted for the PHR (phenol reaction)-negative phenotype in 35 japonica lines (Yu et al., 2008). According to watermelon reference genome, there are eight PPO genes and three of the genes are found in the present narrowed region. Based on the sequence analysis, *ClSC1* encoding polyphenol oxidase enzymes was the casual gene for the black seed coat coloration.

Black seed is attractive with scarlet or canary yellow flesh and is considered as a promising source of the natural melanin, which shows remarkable activities of antioxidant, radio-protective, thermoregulative, antitumor, immunostimulating and anti-inflammatory (Yan et al., 1996; Łopusiewicz, 2018). In the present study, significantly higher melanin content was found in black seed coat compared with light yellow seed coat. Besides, polyphenols content in light yellow seed coat was much higher than in black seed coat (**Figure 2**), which is in harmony with *B. napus* (Ye et al., 2002). For this phenomenon, we hypothesized that a large amount of polyphenols in black seed coat were used to synthesize melanin through the catalyzation of PPO, resulting in the pigment difference in black and yellow light seed coat (Ye et al., 2002). In this study, sequence alignment between light yellow seed coat and black seed coat showed that a single-nucleotide (A) insertion within the *ClSC1* in 9904 (light yellow) resulted in a frameshift mutation and premature stop codon (**Figure 1**). The mutated *ClSC1* encoded a non-functional truncated protein (42 amino acid). In addition, *ClSC1* specifically showed significantly high level of transcript in Handel with black seed coat (**Figure 3**). Moreover, there was significantly higher PPO activity in black seed coat compared with light yellow seed

coat (**Figure 2**). Hence, it is hypothesized that the different expression of *ClSC1* accounts for the different activities of PPO in the seed coat, which resulted in the different amount of melanin forming the black and light yellow seed coat.

In general, we illustrated the inheritance pattern of seed coat color and suggested a causal gene *ClSC1* accounting for the black seed coat coloration in watermelon. Our results further facilitate marker-assisted breeding and provide further evidence to understand the molecular mechanisms in seed coat coloration in watermelon and other crops.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

WL and XL conceived the research and designed the experiments. NH developed the plants population. SZ, PY and CG analyzed data. HG and UM checked the manuscript. BL performed most of the experiment and wrote the manuscript. All authors reviewed and approved this submission.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://.frontiersin.org/articles/10.3389/fpls.2019.01689/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | The analysis of the phylogenetic and conserved domains of the candidate gene. **(A)** The phylogenetic tree of Cla019481 and its homologous proteins. **(B)** The conserved domains were analyzed by online Pfam database.

SUPPLEMENTARY TABLE 1 | The segregation ratio of seed coat color among different populations.

SUPPLEMENTARY TABLE 2 | The information of CAPS and SNP markers on chromosome 3 used for the polymorphic analysis.

SUPPLEMENTARY TABLE 3 | The primers information used for the sequence and expression analysis of the candidate genes.

REFERENCES

Abd el-Hafez, A., Gaafer, A., and Allam, A. (1985). Inheritance of flesh colour, seed coat cracks and total soluble solids in watermelon and their genetic relations. II. Quantitative characters and the association between various characters. *Acta Agron. Acad. Sci. Hung.* 34, 84–89.

Appelhagen, I., Thiedig, K., Nordholt, N., Schmidt, N., Huep, G., Sagasser, M., et al. (2014). Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta* 240 (5), 955–970. doi: 10.1007/s00425-014-2088-0

Baxter, I. R., Young, J. C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., et al. (2005). A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 102 (7), 2649–2654. doi: 10.1073/pnas.0406377102

Bharti, A. K., and Khurana, J. P. (2003). Molecular characterization of transparent testa (tt) mutants of *Arabidopsis thaliana* (ecotype Estland) impaired in flavonoid biosynthetic pathway. *Plant Sci.* 165 (6), 1321–1332. doi: 10.1016/S0168-9452(03)00344-3

Boss, P. K., Gardner, R. C., Janssen, B.-J., and Ross, G. S. (1995). An apple polyphenol oxidase cDNA is up-regulated in wounded tissues. *Plant Mol. Biol.* 27 (2), 429–433. doi: 10.1007/bf00020197

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138 (3), 963–971. doi: 10.1007/bf00020197

Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol.* 122 (2), 403–414. doi: 10.1104/pp.122.2.403

Demeke, T., and Morris, C. (2002). Molecular characterization of wheat polyphenol oxidase (PPO). *Theor. Appl. Genet.* 104 (5), 813–818. doi: 10.1007/s00122-001-0847-3

Demir, I., Mavi, K., and Oztokat, C. (2004). Changes in germination and potential longevity of watermelon (Citrullus lanatus) seeds during development. *New Zealand. J. Crop Hortic. Sci.* 32 (1), 139–145. doi: 10.1080/01140671.2004.9514288

Dry, I. B., and Robinson, S. P. (1994). Molecular cloning and characterisation of grape berry polyphenol oxidase. *Plant Mol. Biol.* 26 (1), 495–502. doi: 10.1007/bf00039560

El-Bramawy, M.E.-H.S., El-Hendawy, S., and Shaban, W. A. (2008). Assessing the suitability of morphological and phenological traits to screen Sesame genotypes for Fusarium wilt and charcoal rot disease resistance. *J. Plant Prot. Res.* 48 (4), 397–410. doi: 10.2478/v10045-008-0049-y

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 (1), 51. doi: 10.1038/ng.2470

Hawkins, L. K., Dane, F., and Kubisiak, T. L. (2001). Molecular markers associated with morphological traits in watermelon. *Hort. Sci.* 36 (7), 1318–1322. doi: 2001; 10.21273/HORTSCI.36.7.1318

He, X., He, Z., Zhang, L., Sun, D., Morris, C., Fuerst, E., et al. (2007). Allelic variation of polyphenol oxidase (PPO) genes located on chromosomes 2A and 2D and development of functional markers for the PPO genes in common wheat. *Theor. Appl. Genet.* 115 (1), 47–58. doi: 10.1007/s00122-007-0539-8

Hunt, M. D., Eannetta, N. T., Yu, H., Newman, S. M., and Steffens, J. C. (1993). cDNA cloning and expression of potato polyphenol oxidase. *Plant Mol. Biol.* 21 (1), 59–68. doi: 10.1007/bf00039618

Kanda, T. (1931). The inheritance of seed-coat coluring in the watermelon. *Jap. J. Genet.* 7, 30–48.

Kong, Q., Yuan, J., Gao, L., Zhao, L., Cheng, F., Huang, Y., et al. (2015). Evaluation of appropriate reference genes for gene expression normalization during watermelon fruit development. *PloS One* 10 (6), e0130865. doi: 10.1007/bf00039618

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

Levi, A., Thomas, C. E., Zhang, X., Joobeur, T., Dean, R. A., Wehner, T. C., et al. (2001). A genetic linkage map for watermelon based on randomly amplified polymorphic DNA markers. *J. Am. Soc. Hortic. Sci.* 126 (6), 730–737. doi: 10.21273/JASHS.126.6.730

Li, X., Westcott, N., Links, M., and Gruber, M. Y. (2010). Seed coat phenolics and the developing silique transcriptome of Brassica carinata. *J. Agric. Food Chem.* 58 (20), 10918–10928. doi: 10.1021/jf102208a

Li, X., Xue, Z.-P., and Zhu, W.-X. (2011). Antioxidant activities and contents of total flavonoids and phenols from different parts of Salvia miltiorrhiza bunge. *Food Sci.* 3. doi: 10.1016/j.carbpol.2011.09.024

Li, B., Lu, X., Dou, J., Aslam, A., Gao, L., Zhao, S., et al. (2018). Construction of a high-density genetic map and mapping of fruit traits in watermelon (Citrullus Lanatus L.) based on whole-genome resequencing. *Int. J. Mol. Sci.* 19 (10), 3268. doi: 10.3390/ijms19103268

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2– ΔΔCT method. *Methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Łopusiewicz, Ł. (2018). Antioxidant, antibacterial properties and the light barrier assessment of raw and purified melanins isolated from Citrullus lanatus (watermelon) seeds. *Herba Pol.* 64 (2), 25–36. doi: 10.2478/hepo-2018-0008

Maina, F. N., and Nyambura, K. (2018). Effect of seed coat colour on the seed quality of water melon (Citrullus Lanatus). *Afr. J. Tech. Vocat. Educ. Training* 3 (1), 195–201.

Marles, M. S., and Gruber, M. Y. (2004). Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the Brassicaceae. *J. Sci. Food Agric.* 84 (3), 251–262.doi: 10.1002/jsfa.1621

Matheis, G., and Whitaker, J. R. (1984). Modification of proteins by polyphenol oxidase and peroxidase and their products. *J. Food Biochem.* 8 (3), 137–162. doi: 10.1111/j.1745-4514.1984.tb00322.x

Mavi, K. (2010). The relationship between seed coat color and seed quality in watermelon Crimson sweet. *Hortic. Sci.* 37 (2), 62–69. doi: 10.17221/53/2009-hortsci

Mayer, A. M., and Harel, E. (1979). Polyphenol oxidases in plants. *Phytochemistry* 18 (2), 193–215. doi: 10.1016/0031-9422(79)80057-6

McKay, J. (1936). Factor interaction in citrullus: seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. *J. Hered.* 27 (3), 110–112. doi: 10.1093/oxfordjournals.jhered.a104182

Mohr, H. (1986). Watermelon breeding. *Breed. Veg. Crops Westport, CT, AVI Publ.* 37–66.

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19), 4321–4326. doi: 10.1093/nar/8.19.4321

Padmaja, L. K., Agarwal, P., Gupta, V., Mukhopadhyay, A., Sodhi, Y. S., Pental, D., et al. (2014). Natural mutations in two homoeologous TT8 genes control yellow seed coat trait in allotetraploid Brassica juncea (AABB). *Theor. Appl. Genet.* 127 (2), 339–347. doi: 10.1007/s00122-013-2222-6

Park, K.-I., and Hoshino, A. (2012). A WD40-repeat protein controls proanthocyanidin and phytomelanin pigmentation in the seed coats of the Japanese morning glory. *J. Plant Physiol.* 169 (5), 523–528. doi: 10.1016/j.jplph.2011.12.001

Poole, C. F. (1941). Inheritance of seed characters in watermelon. *J. Agr. Res.* 63, 433–456.

Paudel, L., Clevenger, J. P., and McGregor, C. (2019). Chromosomal locations and interactions of four loci associated with seed coat color in watermelon. *Front. Plant Sci.* 25, 788. doi: 10.3389/fpls.2019.00788

Qu, C., Fu, F., Lu, K., Zhang, K., Wang, R., Xu, X., et al. (2013). Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded Brassica napus. *J. Exp. Bot.* 64 (10), 2885–2898. doi: 10.1093/jxb/ert148

Ren, R., Ray, R., Li, P., Xu, J., Zhang, M., Liu, G., et al. (2015). Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Mol. Genet. Genomics* 290 (4), 1457–1470. doi: 10.1007/s00438-015-0997-7

Sagasser, M., Lu, G.-H., Hahlbrock, K., and Weisshaar, B. (2002). A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes Dev.* 16 (1), 138–149. doi: 10.1101/gad.212702

Shahar, T., Hennig, N., Gutfinger, T., Hareven, D., and Lifschitz, E. (1992). The tomato 66.3-kD polyphenoloxidase gene: molecular identification and developmental expression. *Plant Cell* 4 (2), 135–147. doi: 10.1105/tpc.4.2.135

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74 (1), 174–183. doi: 10.1111/tpj.12105

Taketa, S., Matsuki, K., Amano, S., Saisho, D., Himi, E., Shitsukawa, N., et al. (2010). Duplicate polyphenol oxidase genes on barley chromosome 2H and their functional differentiation in the phenol reaction of spikes and grains. *J. Exp. Bot.* 61 (14), 3983–3993. doi: 10.1093/jxb/erq211

Tran, L. T., Taylor, J. S., and Constabel, C. P. (2012). The polyphenol oxidase gene family in land plants: lineage-specific duplication and expansion. *BMC Genomics* 13 (1), 395. doi: 10.1186/1471-2164-13-395

Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front. Plant Sci.* 7, 1491. doi: 10.3389/fpls.2016.01491

Wehner, T., and Pitrat, M. (2008). Overview of the genes of watermelon *Cucurbitaceae* EUCARPIA meeting. Pitrat M (ed), 79–89..

Wehner, T. C. (2008). "Watermelon," in *Vegetables I* (New York: Springer), 381–418

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6, 8609. doi: 10.1038/ncomms9609

Whitaker, J. R., and Lee, C. Y. (1995) Recent advances in chemistry of enzymatic browning. *Enzymatic browning and its prevention* 600, 2–7. doi: 10.1021/bk-1995-0600.ch001

Yan, W., Xuehui, L., MaoSu, L., and Rong, Y. P. R. (1996). Study on the structure of severaI natural melanins by FTIR spectrometer. *Anal. Lab.* (6), 6. doi: 10.13595/j.cnki.issn1000-0720.1996.0152

Yan, S., Li, S., Zhai, G., Lu, P., Deng, H., Zhu, S., et al. (2017). Molecular cloning and expression analysis of duplicated polyphenol oxidase genes reveal their functional differentiations in sorghum. *Plant Sci.* 263, 23–30. doi: 10.1016/j.plantsci.2017.07.002

Ye, X.-L., Li, J.-N., Tang, Z.-L., and Chen, L. (2001a). Difference of seedcoat color between black-and yellow-seeded in B. napus with seed development II. Changes of melanin, tyrosine and tyrosinase and their correlation analyses. *Chin. J. Oil Crop Sci.* 3, 39–42.

Ye, X., Li, J., Tang, Z., and Chen, L. (2001b). "Dynamic studies on difference of seedcoat pigments and related characters between black-seeded and yellow-seeded rapeseed lines (*Brassica napus* L.) in the developmental stages of embryo", in: *Proceedings international symposium on rapeseed science*. (New York: Science Press), 73–84.

Ye, X., Li, J., Tang, Z., Liang, Y., and Chen, L. (2001c). Study on seedcoat color and related characters of Brassica napus. *Zuo Wu Xue Bao* 27 (5), 550–556. doi: 10.3321/j.issn:0496-3490.2001.05.002

Ye, X., LI, X., and li, J. (2002). Mechanism of melanin synthesis in seed coat of Brassica napus L. *Acta Agron. Sin.* 28 (5), 638–643. doi: 10.3321/j.issn:0496-3490.2001.05.002

Ye, X., Li, J., Tang, Z., and Chen, L. "Dynamic studies on difference of seedcoat pigments and related characters between black-seeded and yellow-seeded rapeseed lines (Brassica napus L.) in the developmental stages of embryo," in *Proceedings international symposium on rapeseed science* (New York: Science Press), 73–84.

Yu, C., Northwest, A., Leisova, L., Kucera, V., Vyvadilova, M., Ovesna, J., et al. (2007). Assessment of genetic diversity of yellow-seeded rapeseed (Brassica napus L.) accessions by AFLP markers. *Czech J. Genet. Plant Breeding-UZPI* 43 (3), 105–112. doi: 10.17221/2071-CJGPB

Yu, Y., Tang, T., Qian, Q., Wang, Y., Yan, M., Zeng, D., et al. (2008). Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* 20 (11), 2946–2959. doi: 10.1007/s13353-012-0132-y

Yu, C.-Y. (2013). Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. *J. Appl. Genet.* 54 (2), 135–145. doi: 10.1007/s13353-012-0132-y

Zhang, X., Yang, G., Chen, L., Yin, J., Tang, Z., and Li, J. (2006). Physiological differences between yellow-seeded and black-seeded rapeseed (*Brassica napus* L.) with different testa characteristics during artificial ageing. *Seed Sci. Technol.* 34 (2), 373–381. doi: 10.15258/sst.2006.34.2.13

An Improved Melon Reference Genome With Single-Molecule Sequencing Uncovers a Recent Burst of Transposable Elements With Potential Impact on Genes

Raúl Castanera[1†], Valentino Ruggieri[1,2†], Marta Pujol[1,2], Jordi Garcia-Mas[1,2] and Josep M. Casacuberta[1*]*

[1] Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Barcelona, Spain,
[2] Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Genomics and Biotecnology Program, Barcelona, Spain

The published melon (*Cucumis melo* L.) reference genome assembly (v3.6.1) has still 41.6 Mb (Megabases) of sequences unassigned to pseudo-chromosomes and about 57 Mb of gaps. Although different approaches have been undertaken to improve the melon genome assembly in recent years, the high percentage of repeats (~40%) and limitations due to read length have made it difficult to resolve gaps and scaffold's misassignments to pseudomolecules, especially in the heterochromatic regions. Taking advantage of the PacBio single- molecule real-time (SMRT) sequencing technology, an improvement of the melon genome was achieved. About 90% of the gaps were filled and the unassigned sequences were drastically reduced. A lift-over of the latest annotation v4.0 allowed to re-collocate protein-coding genes belonging to the unassigned sequences to the pseudomolecules. A direct proof of the improvement reached in the new melon assembly was highlighted looking at the improved annotation of the transposable element fraction. By screening the new assembly, we discovered many young (inserted less than 2Mya), polymorphic LTR-retrotransposons that were not captured in the previous reference genome. These elements sit mostly in the pericentromeric regions, but some of them are inserted in the upstream region of genes suggesting that they can have regulatory potential. This improved reference genome will provide an invaluable tool for identifying new gene or transposon variants associated with important phenotypes.

Keywords: long-reads, assembly, reference genome, transposable elements, melon

INTRODUCTION

Melon (*Cucumis melo* L.) is one of the most important plant crops, with a worldwide production reaching near 32 million tons in 2017 (http://www.fao.org). A high-quality reference genome assembly of melon was released in 2012 (Garcia-Mas et al., 2012). This assembly was generated using 454 reads and Sanger sequencing of BAC ends and contained up to 375 Mb of sequence

assembled into 1,594 scaffolds, with an N50 of 4.68 Mb. Since then, additional improvements have been performed. In particular, a high-resolution genetic map was used to anchor up to 98.2% of the scaffold assembly to the 2X = 24 melon chromosomes (Argyris et al., 2015), followed by an optical mapping used to improve the orientation of the scaffolds of the previous assembly and accurately define the gap content (Ruggieri et al., 2018). Besides the efforts done to improve the original assembly, the lastest published melon reference genome (v3.6.1) (Ruggieri et al, 2018), still contains up to 19.1% of its sequence in gaps and 41.6 Mb of unassigned sequences (22,123 unassigned contigs out of the 42,067 contigs, grouped as Chr0).

Previous analyses on the melon genome architecture have described that this species contains expanded pericentromeres arising from massive amplification of transposable elements (TEs) in the past 10 million years (Mya) (Morata et al., 2018). TEs tend to accumulate in centromeric and pericentromeric regions due to the preferential insertion of some elements, including some retrotransposon families (Neumann et al., 2011), but also as a consequence of the counter-selection of insertions in genic regions that are more likely to be deleterious (Contreras et al., 2015). Due to the enriched proportion of long, repeated sequences such as Long Terminal Repeat (LTR) retrotransposons and other TEs, plant centromeres, and pericentromeres are difficult to assemble and often contain multiple gaps. Such difficulty arises from the limitation of the short-read sequencing to distinguish between near-identical repeated regions. However, this limitation will also make it difficult to correctly assemble TEs sitting in gene-rich regions. As a consequence, short-read-based assemblies may contain an underestimated transposon content, with elements missing in the pericentromeric regions, but also in the proximity of genes, and potentially impacting on gene regulation or coding capacity. Third generation sequencing offers a great opportunity to improve short-read-based assemblies such as the melon reference genome due to the longer read length, the low systematic bias, high consensus read accuracy, and improved assembling algorithms. In the recent years, many studies have taken advantage of these technologies for improving draft genomes or generating chromosome-level assemblies (Jiao et al., 2017b; Zhang et al., 2019). We describe here a new reference assembly for the cultivated melon (v4.0). This new version benefited from ~50-fold PacBio long reads coupled with a 20-fold Illumina short-reads data, which allowed to improve the characterization and accuracy of several regions of the genome, particularly repetitive regions and centromeric areas.

METHODS

DNA Extraction and Sequencing
Genomic DNA was extracted from the double-haploid line DHL92, the same line sequenced to obtain the previous version of the melon genome, v3.6.1 (Ruggieri et al., 2018), as described by (Doyle, 1991) with minor modifications. Three grams of young leaves were harvested and frozen in liquid nitrogen for tissue homogenization. After isopropanol precipitation, instead of centrifugation, the DNA was recovered by fishing with a little glass hook to avoid fragmentation. We added a purification step using phenol:chloroform:isoamyl alcohol (25:24:1), and resuspended in Milli-Q® water. DNA integrity was evaluated by gel electrophoresis and quantified by Qubit 2.0. DNA was purified with AMPure® PB beads, and length was evaluated with the Fragment Analyzer Femto Pulse (Advanced Analytical Technologies, Inc.). DNA sequencing was performed using Pacific Biosciences (PacBio) RSII technology at the Platform GENTYANE, INRA/UCA (Clermont-Ferrand, France). A total of ~2,5 million PacBio long reads were generated, which corresponds to ~50x coverage of the estimated melon genome.

Genome Assembly
The reads from the PacBio system were assembled using the hierarchical genome-assembly process 4 (HGAP4) pipeline (Pacific Biosciences, SMRT Link Suite 6.0). The principle and workflow of HGAP pipeline consists of different concatenated steps, including (i) the selection of the longest reads as a seeding sequence data set, (ii) the use of each seeding sequence as a reference to recruit shorter reads and preassemble reads through a consensus procedure, (iii) the assembly of the preassembled reads, (iv) the refinement/polishing using all initial read data to generate the final consensus (Chin et al., 2013). In the pre-assembly step, the raw reads were filtered using default settings with read quality (rq) of ≥ 0.65. Then, the assembly step was performed using FALCON in the HGAP4 tool with seed coverage set to 50, "aggressive" option turned on, and minimum accuracy set to 65. The ARROW algorithm was used to polish the genome assembly with default parameters.

Reference-Guided Contig Ordering, Orientation, and Genome Quality Assessment
The contigs produced by the assembly were ordered and oriented based on the latest melon assembly (v3.6.1) with the RaGOO tool, which uses a reference-guided process (Alonge et al., 2019). In order to improve the mappability of PacBio contigs, a polishing step was previously performed on the v3.6.1 assembly using raw PacBio reads. With this aim, the ARROW pipeline in the SMRT Link suite (resequencing pipeline) was used with default parameters, just superimposing the minimum number of reads to call variant ≥ 15.

RaGOO is an open-source tool, implemented as a python command-line utility, which internally invokes Minimap2 (Li, 2018). Default parameters were used with k-mer size and window size both set to 19 bp. Any alignment shorter than 1 kbp in length was removed. As reported by the author, to cluster contigs, the tool assigns each contig to the reference chromosome which it covers the most. Subsequently, for each pseudomolecule group, the contigs in that group are ordered and oriented relative to each other by examining the longest (primary) alignment. Ordering is then achieved by sorting these primary alignments. To produce pseudomolecules, the contigs are concatenated, with

padding of 1,000 "N" characters placed between contigs. Finally, the new consensus sequences were polished with 20× Illumina paired-end reads (2 × 150 bp). Reads were aligned to the assembly using BWA-MEM (Li and Durbin, 2009). Sequence error correction was performed with the Pilon pipeline (Walker et al., 2014). The completeness of the final assembly was evaluated with BUSCO (version 3) (Simão et al., 2015) using the conserved plant genes as database (Eudicotyledons *odb10**). Comparative analysis and synteny between v3.6.1 and v4.0 assemblies were performed using MAUVE (Darling et al., 2004) and SyMAP v4.2 (Soderlund et al., 2011).

Genome Annotation

The genome annotation was performed by transferring through a liftover process the latest published gene models (Ruggieri et al., 2018) to the new PacBio-based genome assembly using Maker v2 program (Campbell et al., 2014). The parameters used in the configuration file were the following: est_forward = 1, est2genome = 1, split_hit = 20000, min_intron = 20, single_exon = 1, single_length = 149, correct_est_fusion = 1. In case of a gene mapping on different positions of the genome, only the match with the highest Maker score was retained. Gene ontology (GO) enrichment analysis was carried out using GOATOOLS (Klopfenstein et al., 2018)

Annotation of Transposable Elements

Transposable elements were detected in the new genome assembly using TEdenovo pipeline from the REPET package (Flutre et al., 2011), excluding the structural search. Consensus sequences representing each TE family were classified into TE orders using PASTEC (Hoede et al., 2014) and annotation of TE copies was carried out by TEannot using two iterations. After the first TEannot run, only consensus sequences that had a full-length match in the genome were retained. A second iteration of TEannot using these consensuses was used to obtain the final annotation. We used blastx (Repbase peptide database (Bao et al., 2015), cut off e-value = e-5) to identify TIR-TE copies that retained coding potential.

Specific Annotation of LTR-Retrotransposons

LTR-retrotransposon candidates were detected by a structural approach using LTRharvest (Ellinghaus et al., 2008). Every element was translated to the six possible frames and scanned for LTR-retrotransposon-specific domains using hmmscan (Eddy, 2011). Elements without coding potential were filtered out, and the remaining elements were classified into Copia and Gypsy superfamilies based on the order of the internal coding domains, as defined by (Xiong and Eickbush, 1990). Elements lacking one or more domains were tagged as "unclassified".

Insertion Age of LTR-Retrotransposons

The LTR regions of every coding element were extracted and aligned with MUSCLE (Edgar, 2004). Kimura 2P distance of every aligned LTR pair was calculated and used to estimate insertion age as previously reported (SanMiguel et al., 1998),

using the *Arabidopsis* mutation rate of 7×10^{-9} nucleotides per site per year (Ossowski et al., 2010).

Identification of Polymorphic LTR-Retrotransposons

Resequencing short-read data from six melon varieties (CV, IRAK, PI 161375, Trigonous, Calcuta, and Vedrantais) were previously available (Garcia-Mas et al., 2012; Sanseverino et al., 2015). Trimming and adapter removal was performed with AdapterRemoval (Lindgreen, 2012). Clean reads were mapped to the v4.0 assembly using BWA-MEM (Li and Durbin, 2009). PINDEL (Ye et al., 2009) was run on mapping files to identify deletions in re-sequenced varieties, using the following parameters: Minimun mapping quality = 35, minimum number of supporting reads for calling a deletion = 5. A polymorphic insertion was scored when the deletion and reference element displayed a reciprocal intersect of 90% of the length.

RESULTS

Genome Assembly Workflow

The approach followed relies on a combination of different pipelines and resources as highlighted in **Supplementary Figure S1**. The first step took as input four PacBio runs, which yielded about 21 Gigabases of sequence (corresponding approximately to a 50x melon genome coverage) with an average read length of 8 kbp and an N50 of about 15 kbp (**Supplementary Table S1**). At the pre-assembly stage, 1,499,406 seed reads were selected with an average length of 12 Kbp. The seed reads were used to produce about 1,469,624 pre-assembled reads (**Supplementary Table S1**). The final consensus assembly yielded 1,178 contigs with a N50 of 714 kbp and a total genome size of 357.64 Mbp. The final mean coverage obtained and the realigned subread concordance are illustrated in **Supplementary Figure S2**.

Pseudomolecule Construction

Following a reference-guided process, the contigs produced by the assembly were ordered and oriented based on the latest melon assembly (v3.6.1) (**Supplementary Figure S1**). In order to improve the mappability of the PacBio contigs on the genome, a polishing/correction step using the complete set of PacBio reads was undertaken. A total of 648,906 variants (271,290 deletions, 293,163 insertions and 84,453 substitutions) on the published v3.6.1 genome were corrected, leading to 1% improvement in mapping of the PacBio contigs. During pseudomolecule construction, we could assign 96% of the contigs to the 12 melon chromosomes, leaving only 44 unassigned short contigs (average length of 7.2 kbp).

Further Polishing of Pseudomolecules

The last step of the workflow was aimed to correct/polish the PacBio assembly using 20-fold Illumina reads from a previous study (Sanseverino et al., 2015). A total of 169,279 variants
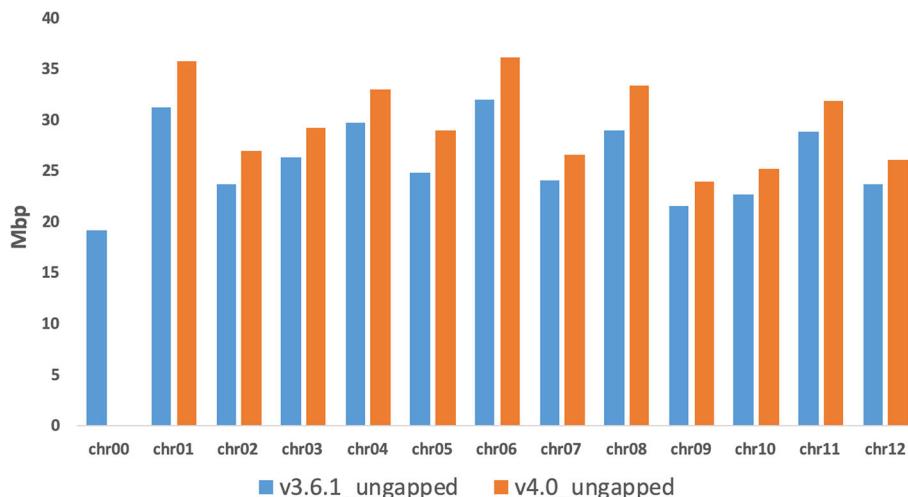
FIGURE 1 | Comparison of the chromosomes length (ungapped) between the v3.6.1 and the v4.0 genome assemblies.

(117,320 insertions, 29,353 deletions, 22,606 substitutions), representing less than 0.1% of the total genome size in length, were detected and included in the final genome v4.0. After error correction using short sequence reads, the total size of melon pseudomolecules is 358 Mb. The new reference assembly contains 1,169 artificial gaps (strings of 1,000 Ns) and has a much higher contiguity than the previously published short-read genome assembly DHL92 v3.6.1 (contig N50 improved from 26.1 kb to 714 kb; contig number improved from 42,067 to 1,178). **Figure 1** shows an improvement of the v4.0 genome assembly in terms of length increase of each chromosome and reduction of unassigned contigs in Chr0 reduction when compared with v3.6.1 assembly. These results highlight the increase of the pseudomolecules sizes of about 40 Mbp (about 20 Mbp already present in Chr0 and 20 Mbp of completely new sequence) in this new assembly, which corresponds approximately to 11% of the total genome length. **Supplementary Table S2** reports the anchoring of 21,283 unassigned contigs of the v3.6.1 Chr0 (96.2%) on the new PacBio melon assembly. A synteny analysis of v3.6.1 and v4.0 assemblies showed a high degree of correspondence across all chromosomes, with short re-oriented or reordered blocks on all chromosomes except in Chr3, Chr8, and Chr11 (**Supplementary Figures S3** and **S4**). In terms of block relocation among chromosomes noteworthy changes were detected between Chr02, Chr11, and Chr12 of v4.0 assembly and Chr05, Chr06, Chr08, and Chr10 of the v3.6.1 assembly, respectively (**Supplementary Figure S3**).

In order to assess the level of completeness of the new assembly with respect to the gene content, we performed a BUSCO analysis. We obtained 94.8% of complete and 1.7% of fragmented BUSCOs at the genome level and a 91.1% of complete and 2.4% fragmented BUSCOs at the gene model level. The observed values are comparable to the ones reported in the v3.6.1 genome assembly, suggesting that the previous

assembly has captured most of the gene information. To maintain gene models and names, the current annotation was transferred to the v4.0 PacBio assembly through a liftover process. We successfully moved 28,299 out of 29,980 gene models to the new genome assembly. The 5% (1,618) of transcripts that did not pass the MAKER's thresholds to define a proper gene model mainly consist of proteins with unknown function (651), transposons (79), and girdin-like proteins (39). In terms of distribution, about 22% of them (374) were from the unassembled contigs in Chr0. This failure could be due to the fact that part of these genes, especially those with unknown functions, may represent false or partial gene models in the previous genome annotation. The re-arrangements of some contigs in the new assembly may also be in part responsible of this discrepancy. A complete list of these genes is provided in **Supplementary Table S3**.

Assembly v4.0 Captures a Larger Fraction of Repetitive Elements

We used the TEdenovo pipeline from the REPET package to identify TE sequences from the melon v4.0 assembly and to build TE consensuses. Subsequently, two iterations of TEannot were run to annotate TE sequences. Transposons cover 45.2% of the new genome assembly (excluding unclassified sequences), in comparison to 35.7% found in the v3.6.1 (Morata et al., 2018). Similarly to what was found in the v3.6.1 genome assembly, LTR-retrotransposons represented the largest fraction of TEs in the v4.0 assembly, followed by Terminal Inverted Repeats (TIRs) containing TEs (TIR-TEs) (**Table 1**). The number of LTR-retrotransposons is higher for v4.0 but the genome fraction that LTR-retrotransposons account for in the two assemblies is similar. On the contrary, we observed a drastic increase in the amount of annotated TIR-TEs and the genome fraction they account for in v4.0 as compared with v3.6.1 (14.97% and 7.11% respectively). We tested for TIR-TE copies that retained coding

TABLE 1 | Comparison of the transposable element (TE) annotation based on the v3.6.1 and v4.0 assemblies.

TE order	Acronym	v3.6.1		v4.0	
		Copies	Genome fraction (%)	Copies	Genome fraction (%)
LTR	RLX	74,161	23.44	136,761	23.81
LINE	RIX	11,913	2.64	15,067	1.80
SINE	RSX	391	0.04	746	0.04
DIRS	RYX	4,212	1.65	17,890	4.39
TIR	DTX	21,383	7.11	92,819	14.97
Helitron	DHX	1,699	0.3	5,637	0.45
Others		912	0.48	823	0.07
TOTAL			35.66		45.53

The TE orders are referred to as follows: LTR-retrotransposons (LTR), Long Interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), DIRS retrotransposons (DIRS), TIR-TEs (TIR), and Helitrons.

potential and found that in both cases the vast majority of the annotated elements were non-coding (89.1% in v3.6.1 and 96.1% in v4.0). The v4.0 assembly has 851 more coding TIR-TEs as compared with v3.6.1. However, the main difference between the TIR-TE fraction of both assemblies is explained by the differential amount of non-coding elements. The size distribution of TIR-TEs (**Supplementary Figure S5**) also

TABLE 2 | Annotation of full-length LTR-retrotransposons. Number of full-length retrotransposon copies belonging to Gypsy, Copia, and unclassified superfamilies in the published v3.6.1 and the v4.0 genome assemblies.

Superfamily	v3.6.1	v4.0
Gypsy	815	1,526
Copia	1,067	1,427
Unclassified	1,358	1,607
TOTAL	3,240	4,560

supports this result, as the biggest differences between the two annotations are found for sizes between 100–500 bp, which are compatible with the length of MITEs and partial TE copies. These differences can be attributed in part to different annotation thresholds. Indeed, the peak found at 100 bp in v4.0 absent in v3.6.1 reflects a difference in the annotation approach (minimum annotation size = 200bp in v3.6.1). However, v4.0 contains more TIR-TEs elements of all sizes, and in particular of elements with a size shorter than 1,000 nt that probably represent truncated TIR-TEs and MITEs, which could be the result of a more complete representation of repetitive sequences in the assembly.

Assembly v4.0 Contains Many Young LTR-Retrotransposons Missing In v3.6.1

Among the different classes of TEs that populate plant genomes, young LTR-retrotransposons are the most difficult to assemble due to their length and high similarity between copies. LTR-retrotransposons are frequently abundant and show a high level of polymorphism in varieties and individuals, which make them important targets of study. In order to annotate these elements in the v4.0 assembly and compare the LTR-retrotransposon content with that of v3.6.1, we used a structural and homology-based approach to identify and to date LTR-retrotransposon insertions. Using this approach, we annotated 1,320 full-length elements more in v4.0 than in v3.6.1, which represents an increase of 40% (**Table 2**). An important fraction of the new LTR-retrotransposons belongs to the Gypsy superfamily, but the v4.0 assembly also contains more Copia LTR-retrotransposons than the v3.6.1. We dated the insertion of all full-length LTR-retrotransposons, and the results showed that the vast majority of newly assembled elements in v4.0 are very young, with estimated insertion times from 0 to 2 Mya (**Figure 2**). The

FIGURE 2 | Distribution of insertion age of Gypsy, Copia, and unclassified LTR-retrotransposons annotated on the genome assemblies v3.6.1 and v4.0.

distribution of v4.0-specific LTR-retrotransposons followed the general distribution profile of all the annotated TEs, with an accumulation along the centromeres and pericentromeres in all the chromosomes (**Figure 3**), their abundance decreasing in gene-rich regions.

Young LTR-Retrotransposons Are Highly Polymorphic and Have a Potential Impact on Genes

In order to determine to what extent the new assembled elements missing in v3.6.1 had a potential impact on genes, we analyzed in detail all the young, full-length LTR-retrotransposons (0-2 Mya). V3.6.1 contains 443 of these elements, whereas v4.0 contains up to 1,523. Using resequencing data from six varieties and the new v4.0 assembled genome as a reference, we were able to determine the level of polymorphism of these elements. More than half (777) of these young elements were predicted to be absent in at least one of the six varieties. The newly discovered young LRT-retrotransposons in the genome assembly v4.0 were further studied for their potential impact on genes. We found that 116 out of the 1,523 were located in the close upstream regions of annotated genes (< 1,000 bp, **Supplementary Table S4**), and

therefore may be affecting the promoters of such genes. In addition, almost 60% of these elements (69) were predicted to be polymorphic in the six varieties analyzed (**Supplementary Table S5**). An example is the polymorphic Gypsy LTR-retrotransposon inserted into the promoter of an AGAMOUS MADS box transcription factor (MELO3C000260, **Figure 4**). This element is young (0.2 Mya), and predicted to be absent in 4 out of the 6 re-sequenced varieties. This Gypsy element could not be properly assembled in v3.6.1, which shows several gaps in the corresponding region upstream the AGAMOUS gene. In addition, in the v3.6.1 assembly the gene was located in the artificial Chr0, which contained the unassembled contigs, whereas in v4.0 assembly we could place it in Chr11 at the position 23,043,868 bp-23,044,427 bp. A manual inspection of this region allowed the correction of the AGAMOUS MADS box transcription factor gene by combining both MELO3C0002360 and MELO3C019694 (**Supplementary Figure S6**). Other examples of genes carrying a newly assembled LTR-retrotransposon absent in v3.6.1 are a TMV resistance protein N-like (MELO3C021852.2) and a UV radiation resistance-associated protein (MELO3C020442), among others (**Figure 4**, **Supplementary Table S4**). A gene ontology enrichment analysis
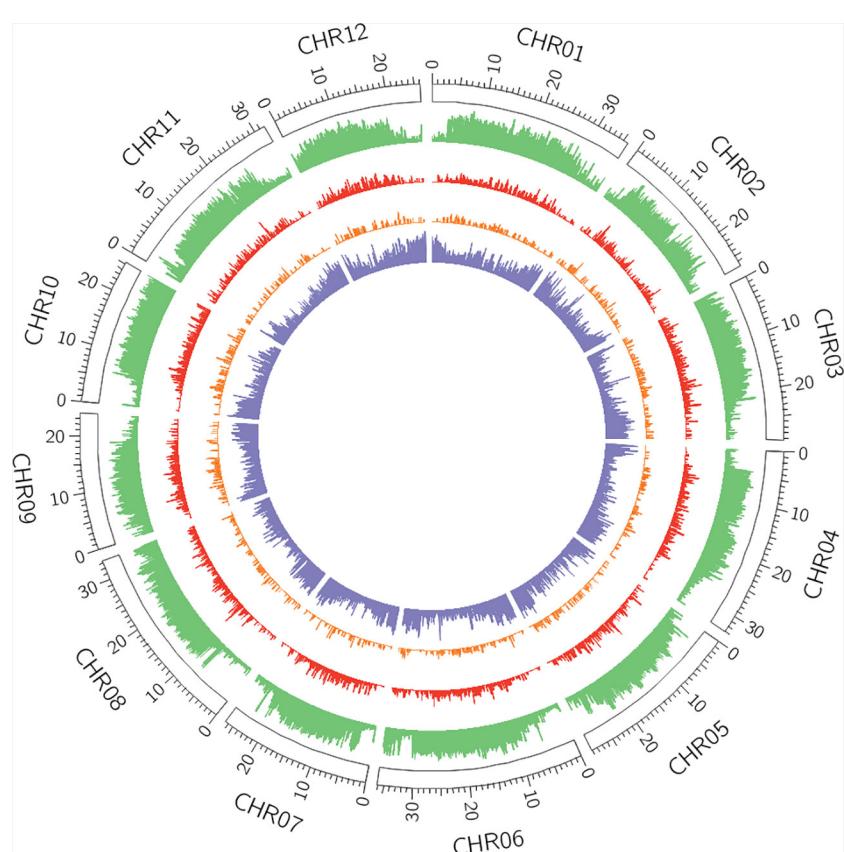
FIGURE 3 | Distribution of transposable elements (TEs) and genes across v4.0 pseudomolecules. In green, density of REPET features per window (6,000 windows in total). In red, density of full-length Long Terminal Repeat (LTR)-retrotransposons annotated in the v4.0 that were absent in the v3.6.1 assembly. In orange, density of polymorphic LTR-retrotransposons with insertion time below 2 Mya. In purple, gene density.

FIGURE 4 | Example of LTR-retrotransposon insertion (red boxes) in the proximal upstream region of genes (blue boxes) annotated in the V4.0 assembly and that corresponded to a gap in the v3.6.1 assembly.

found no enriched terms within the functional annotation of the 116 genes.

DISCUSSION

An Improved Melon Reference Genome Assembly Produced Using Long-Read Sequencing

A high-quality and accurate reference genome represents a relevant resource for basic and applied research including functional genetics, comparative genomics, and population genetics (Kingan et al., 2019). Indeed, many reference genomes for crop plants have been generated over the past decade, even though most of them are often fragmented and missing complex repeat regions (Jiao et al., 2017b). Melon is a widely cultivated crop in the world, and its reference genome was first published in 2012 (Garcia-Mas et al., 2012). This reference genome has been improved over the time (Argyris et al., 2015; Ruggieri et al., 2018), and consists of 42,067 small contigs, assembled in 13 scaffolds. Some of the contigs were still arbitrarily ordered and oriented, which complicated the analysis of some individual loci. In addition, the last published version of the assembly, version v3.6.1, also contains a high number of short gaps, frequently found in intergenic regions and often close to genes. These drawbacks are a limitation for genotype to phenotype analyses, as gaps may contain sequence variability that cannot be used for GWAS or fine-mapping studies, and can also contain candidate genes that cannot be associated to the trait. On the other hand, finding significant SNPs scattered across unassigned scaffolds

can complicate the interpretation of GWAS. All these limitations of incomplete assemblies for genotype to phenotype studies have been previously highlighted (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018; Benevenuto et al., 2019). Here, we combined PacBio (50x) with Illumina (20x) reads to improve the genome assembly v3.6.1 of the melon reference genome DHL92. The use of 2nd generation Illumina sequencing technology to correct PacBio long reads is reported to be an efficient and cost effective way to improve a genome assembly (Mahmoud et al., 2019). This integrated workflow produced an improvement of the genome assembly both in terms of new sequence gained (20 Mbp) and inclusion of previous unassigned contigs (20 Mb). In addition, short blocks were reoriented or reordered within and across chromosomes. The structure of the genome is therefore improved in assembly version v4.0 presented here. The assessment of genome completeness and sequence accuracy of the v4.0 assembly was performed using a set of "Eudicotyledon" conserved genes. This analysis indicated that the main assembly improvements occurred in non-genic regions, in line with what has already been reported for other genomes (Jiao et al., 2017a). The number of gaps was reduced from 44,650 in the genome version v3.6.1 to 1,169 in v4.0 (**Supplementary Figure S7**). The remaining gaps probably result from the presence of very complex regions in the genome that will need further efforts to be solved.

V4.0 Assembly Uncovers a Burst of Young LTR-Retrotransposons

The new v4.0 assembly of the melon genome contains up to 10% more TE content than v3.6.1, a difference that can be explained

mainly by a better capture of coding and non-coding TIR-TEs, as well as a better identification of young LTR-retrotransposons. Non-coding TIR-TEs such as MITEs have been described to be involved in gene regulation through the amplification of transcription factor binding sites (TFBS) (Hénaff et al., 2014; Morata et al., 2018). Thus, our new dataset represents a significant improvement that can be used to assess the functional impact of these elements with a much better precision. Besides the importance of TIR-TEs, LTR-retrotransposons are the most interesting TEs due to their high abundance and their potential impact on genes. In this v4.0 assembly, the LTR-retrotransposon content is similar to that of v3.6.1 in percentage of genome fraction. However, we found a large difference in the content of full-length and young elements. Full LTR-retrotransposons are difficult to annotate with approaches that use genome self-comparison followed by RepeatMasker annotation (i.e., as REPET does). This approach can be effectively used to identify truncated and degenerated copies, but often leads to the fragmentation of long intact elements. To overcome this problem, a structural detection (LTRharvest) followed by a homology-based approach was used to identify full-length elements with coding potential in both v3.6.1 and v4.0 assemblies. Using the same annotation pipeline, we were able to identify up to 40% more full-length LTR-retrotransposons in the new v4.0 assembly, an important fraction of which are located in centromeric and pericentromeric regions. It is well known that Gypsy elements tend to integrate in such regions, which are highly repetitive and difficult to assemble. In this sense, this result evidences that v4.0 assembly captures a much larger fraction of the pericentromeres than v3.6.1 due to the improved assembly of LTR-retrotransposons, especially the younger ones. Our results evidence that a recent (less than 2 Mya) burst of LTR-retrotransposons occurred in the melon genome, which was overlooked in previous analysis due to the incompleteness of the reference assembly. Based on our results, we warn that comparisons of LTR-retrotransposon content and distribution between genome assemblies of very different quality could be strongly biased and should be carefully discussed.

Impact of Young LTR-Retrotransposons on Genes

The genome of melon has been described to have recently expanded pericentromeric regions resulting from a massive TE amplification (Garcia-Mas et al., 2012). The number of young LTR-retrotransposons found in these regions supports the hypothesis that the pericentromeric expansion of melon occurred after the split with cucumber, which was dated about 10 Mya (Sebastian et al., 2010). Here, we have annotated a higher amount of young LTR-retrotransposons (< 2 Mya) located in pericentromeric regions providing an additional support to the hypothesis of the expansion of these regions through the accumulation of LTR-retrotransposon insertions. In addition to the important number of previously unassembled LTR-retrotransposons sitting in the pericentromeric regions, the v4.0 assembly also contains a high number of new LTR-

retrotransposons in gene-rich regions. The detection of recent LTR-retrotransposon insertions at close proximity of genes (< 1 kb away) indicated potential to alter or regulate gene expression. In this study, 60% of these LTR-retrotransposons were first found to be polymorphic in the six melon varieties, providing probable association with phenotypic variation in melon species. Further studies should be addressed to demonstrate this hypothesis. One of these young Gypsy LTR-transposons is inserted into the promoter of the AGAMOUS MADS box transcription factor MELO3C019694, which was miss-annotated in assembly v3.6.1 (**Supplementary Figure S6**). Recently, MELO3C019694 has been suggested as the candidate gene for the presence of sutures trait after performing GWAS and bi-parental mapping experiments (Zhao et al., 2019), and the orthologous SHP1 and SHP2 in Arabidopsis regulates pod dehiscence in this plant (Liljegren et al., 2000). The insertion of a Gypsy element in the promoter of MELO3C019694 will have to be tested in a wide collection of non- and sutured accessions.

CONCLUSION

We present here a new assembly of the melon genome, based on a combination of PacBio and Illumina sequencing, with an improved sequence content and continuity with respect to the previous published assembly version. The v4.0 genome assembly enables identification of important recent LTR-retrotransposon insertions at genes and their polymorphism among melon varieties. These insertions may affect the coding capacity or the expression of melon genes and may be linked to phenotypic variability in agronomic traits, as for example, the presence of sutures in the fruit.

DATA AVAILABILITY STATEMENT

The raw sequencing data and the assembly are available at the European Nucleotide Archive, ENA PRJEB34181. The assembly and the gene and TE annotations are available at the Melonomics database (www.melonomics.net).

AUTHOR CONTRIBUTIONS

JG-M and JC conceived the project. MP obtained the DNA. VR and RC obtained and analyzed the data. VR, RC, JG-M, and JC drafted the manuscript. All authors revised and approved the manuscript.

FUNDING

Programme for Centres of Excellence in R&D 2016-2010 (SEV-2015-0533) and the CERCA Programme/Generalitat de Catalunya to both groups. RC was recipient of a Juan de la Cierva Postdoctoral fellowship from the Spanish Ministerio de Economia y Competitividad.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01815/full#supplementary-material

REFERENCES

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., et al. (2019). Fast and accurate reference-guided scaffolding of draft genomes. *BioRxiv Genome Biol.* 20, 224. doi: 10.1186/s13059-019-1829-6

Argyris, J. M., Ruiz-Herrera, A., Madriz-Masis, P., Sanseverino, W., Morata, J., Pujol, M., et al. (2015). Use of targeted SNP selection for an improved anchoring of the melon (Cucumis melo L.) scaffold genome assembly. *BMC Genomics* 16, 4. doi: 10.1186/s12864-014-1196-3

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11. doi: 10.1186/s13100-015-0041-9

Benevenuto, J., Ferrão, L. F. V., Amadeu, R. R., and Munoz, P. (2019). How can a high-quality genome assembly help plant breeders? *Gigascience* 8 (6), giz068. doi: 10.1093/gigascience/giz068

Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinf.* 48, 4.11.1–4.1139. doi: 10.1002/0471250953.bi0411s48

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Contreras, B., Vives, C., Castells, R., and Casacuberta, J. M. (2015). "The impact of transposable elements in the evolution of plant genomes: from selfish elements to key players," in *Evolutionary Biology: Biodiversification from Genotype to Phenotype*. Ed. P. Pontarotti (Cham: Springer International Publishing), 93–105. doi: 10.1007/978-3-319-19932-0_6

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Doyle, J. (1991). "DNA protocols for plants," in *Molecular techniques in taxonomy*. Eds. G. M. Hewitt, A. W. B. Johnston and J. P. W. Young (Berlin, Heidelberg: Springer Berlin Heidelberg), 283–293. doi: 10.1007/978-3-642-83962-7_18

Eddy, S. R. (2011). Accelerated profile HMM searches. *PloS Comput. Biol.* 7, e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinf.* 9, 18. doi: 10.1186/1471-2105-9-18

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in *de novo* annotation approaches. *PloS One* 6, e16526. doi: 10.1371/journal.pone.0016526

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., et al. (2012). The genome of melon (Cucumis melo L.). *Proc. Natl. Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109

Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C., et al. (2014). Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. *Plant J.* 77, 852–862. doi: 10.1111/tpj.12434

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTEC: an automatic transposable element classification tool. *PloS One* 9, e91929. doi: 10.1371/journal.pone.0091929

International Wheat Genome Sequencing Consortium (IWGSC) and IWGSC RefSeq principal investigators:, Appels, R., Eversole, K., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191. doi: 10.1126/science.aar7191

Jiao, W.-B., Accinelli, G. G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., et al. (2017a). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Res.* 27, 778–786. doi: 10.1101/gr.213652.116

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017b). Improved maize reference genome with single-molecule technologies. *Nature* 546, 524–527. doi: 10.1038/nature22971

Kingan, S. B., Heaton, H., Cudini, J., Lambert, C. C., Baybayan, P., Galvin, B. D., et al. (2019). A high-quality de novo genome assembly from a single mosquito using PacBio sequencing. *Genes (Basel)* 10 (1), 62. doi: 10.3390/genes10010062

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* 8, 10872. doi: 10.1038/s41598-018-28948-z

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

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* 404, 766–770. doi: 10.1038/35008089

Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5, 337. doi: 10.1186/1756-0500-5-337

Mahmoud, M., Zywicki, M., Twardowski, T., and Karlowski, W. M. (2019). Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics* 111, 43–49. doi: 10.1016/j.ygeno.2017.12.011

Morata, J., Tormo, M., Alexiou, K. G., Vives, C., Ramos-Onsins, S. E., Garcia-Mas, J., et al. (2018). The evolutionary consequences of transposon-related pericentromer expansion in melon. *Genome Biol. Evol.* 10, 1584–1595. doi: 10.1093/gbe/evy115

Neumann, P., Navrátilová, A., Koblížková, A., Kejnovský, E., Hřibová, E., Hobza, R., et al. (2011). Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob. DNA* 2, 4. doi: 10.1186/1759-8753-2-4

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science* 327, 92–94. doi: 10.1126/science.1180677

Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., et al. (2018). An improved assembly and annotation of the melon (Cucumis melo L.) reference genome. *Sci. Rep.* 8, 8088. doi: 10.1038/s41598-018-26416-2

SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y., and Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20, 43–45. doi: 10.1038/1695

Sanseverino, W., Hénaff, E., Vives, C., Pinosio, S., Burgos-Paz, W., Morgante, M., et al. (2015). Transposon insertions, structural variations, and snps contribute to the evolution of the melon genome. *Mol. Biol. Evol.* 32, 2760–2774. doi: 10.1093/molbev/msv152

Sebastian, P., Schaefer, H., Telford, I. R. H., and Renner, S. S. (2010). Cucumber (Cucumis sativus) and melon (C. melo) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14269–14273. doi: 10.1073/pnas.1005338107

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

Soderlund, C., Bomhoff, M., and Nelson, W. M. (2011). SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Research* 39 (10), 1 May (2011), e68 doi: 10.1093/nar/gkr123

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection

and genome assembly improvement. *PloS One* 9, e112963. doi: 10.1371/journal.pone.0112963

Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362. doi: 10.1002/j.1460-2075.1990.tb07536.x

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871. doi: 10.1093/bioinformatics/btp394

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* 10, 1494. doi: 10.1038/s41467-019-09518-x

Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., et al. (2019). A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* 51, 1607–1615. doi: 10.1038/s41588-019-0522-8

The MADS-Box Gene *CsSHP* Participates in Fruit Maturation and Floral Organ Development in Cucumber

Zhihua Cheng[1], Shibin Zhuo[1], Xiaofeng Liu[1], Gen Che[1], Zhongyi Wang[1], Ran Gu[1], Junjun Shen[1], Weiyuan Song[1], Zhaoyang Zhou[1], Deguo Han[2*] and Xiaolan Zhang[1*]

[1] State Key Laboratories of Agrobiotechnology, Beijing Key Laboratory of Growth and Developmental Regulation for Protected Vegetable Crops, MOE Joint Laboratory for International Cooperation in Crop Molecular Breeding, China Agricultural University, Beijing, China, [2] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of Northeast Region, Ministry of Agriculture, College of Horticulture & Landscape Architecture, Northeast Agricultural University, Harbin, China

Cucumber is an important vegetable crop bearing fleshy pepo fruit harvested immature. Fruits left unpicked in time during summer production, as well as unfavorable environmental conditions during post-harvest shelf, will cause cucumber fruits to turn yellow and ripen, and thus impair the market value. Identification of maturity-related genes is of great agricultural and economic importance for cucumber production. Here, we isolated and characterized a MADS-box gene, *Cucumis sativus SHATTERPROOF* (*CsSHP*) in cucumber. Expression analysis indicated that *CsSHP* was specifically enriched in reproductive organs including stamens and carpels. Ectopic expression of *CsSHP* was unable to rescue the indehiscence silique phenotype of *shp1 shp2* mutant plant in *Arabidopsis*. Instead, overexpression of *CsSHP* resulted in early flowering, precocious phenotypes, and capelloid organs in wild-type *Arabidopsis*. Biochemical analysis indicated that CsSHP directly interacted with cucumber SEPALLATA (SEP) proteins. *CsSHP* expression increased significantly during the yellowing stage of cucumber ripening, and was induced by exogenous application of abscisic acid (ABA). Therefore, CsSHP may participate in fruit maturation through the ABA pathway and floral organ specification *via* interaction with CsSEPs to form protein complex in cucumber.

Keywords: cucumber, *Cucumis sativus SHATTERPROOF*, floral organ identity, fruit maturation, abscisic acid

INTRODUCTION

The fruit is a major evolutionary success in angiosperms which is essential for plant sexual reproduction and environmental adaptation (Fourquin and Ferrandiz, 2012). The main functions of fruits are to protect and nourish the developing seeds, and to act as a seed dispersal agent (van der Knaap et al., 2014). Angiosperms have evolved different types of fruit to meet the need of diverse dispersal strategies, such as the dry dehiscent fruit in *Arabidopsis thaliana* opens through dehiscence zones to release seeds, whereas the fleshy fruit in tomato attracts frugivorous animals to disperse

seeds by means of bright color and pleasant aromas upon ripening (Roeder and Yanofsky, 2006; Ferrandiz and Fourquin, 2014). Moreover, fruits are the commercial organs for many agricultural crops and play important roles for human diet and health, thus the fruit has been under strong selective pressure during crop domestication (van der Knaap et al., 2014).

The fruit is generally developed from the ovary, which is an important component of the gynoecium. Gynoecium is located in the center of the flower, and surrounded sequentially as whorls by stamens, petals, and sepals, respectively (Robles and Pelaz, 2005). According to the ABC model, the sepal is specified by A function genes, the petal is determined by A+B function genes, the stamen is controlled by the B+C function genes, and the carpel is specified by the C class of genes (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994; Fourquin and Ferrandiz, 2012). In *Arabidopsis,* the *AGAMOUS* (*AG*) gene is the C class of gene that determines the carpel identity, specifies stamen identity with B-function genes, inhibits A-function genes and controls floral meristem determinacy (Bowman et al., 1989; Bowman et al., 1991; Mizukami and Ma, 1992; Mizukami and Ma, 1995). Subsequent studies showed that *SEPALLATA* (*SEP*) genes, expressing in four floral whorls, act as co-factors with ABC homeotic genes in specifying all types of floral organs (Theissen and Saedler, 2001; Favaro et al., 2003; Robles and Pelaz, 2005; Ruelens et al., 2017). Strikingly, all the genes involved in floral organ specification in *Arabidopsis* are from MADS-box transcription factor family excepting the gene *APETALA2* (Jofuku et al., 1994; Dreni and Kater, 2014). MADS-box genes are reported to be the key players in organ morphogenesis throughout the plant life cycle, with a typical MADS domain and a K-box domain in their protein structure (Smaczniak et al., 2012).

There are two MADS-box genes, the *SHATTERPROOF1* (*SHP1*) and *SHP2*, in the AG subfamily of *Arabidopsis* acting as the major regulators directing dehiscence zone differentiation and stimulating lignification of adjacent cells in siliques (Liljegren et al., 2000). In the double mutant *shp1 shp2* plant, the mature siliques were unable to dehisce due to failure of dehiscence zone formation (Liljegren et al., 2000). Constitutive expression of *SHP* genes led to small fruits with overlignified valves (Liljegren et al., 2000). In other dry dehiscent fruits such as *Nicotiana benthamiana* (*NbSHP*), *Glycine max* (*GmAGL1*), and *Medicago*, the SHP gene promotes lignin accumulation in fruit pods to ensure cracking upon maturation (Liljegren et al., 2000; Fourquin and Ferrandiz, 2012; Fourquin et al., 2013; Chi et al., 2017). In the fleshy berry fruit tomato (Giovannoni, 2007), the *SHP1/2* ortholog *TAGL1* participates in fruit expansion and promotes fruit ripening. Knockdown of *TAGL1* resulted in yellow orange fruit with decreased carotenoids and thin pericarps. Overexpression of *TAGL1* led to enlarged sepals and overaccumulation of lycopene, supporting the roles of *TAGL1* in fruit ripening (Itkin et al., 2009; Vrebalov et al., 2009; Gimenez et al., 2010). Similarly, in the fleshy false fruit of strawberry, *FaSHP* was shown to promote maturation as well (Itkin et al., 2009; Vrebalov et al., 2009; Daminato et al., 2013). In citrus, CsMADS6 (the ortholog of SHP1/2) positively modulated

carotenoid metabolism by directly regulating the expression of carotenogenic genes, suggesting an active role in fruit ripening (Lu et al., 2018). In addition to the functions in fruit opening and ripening, *SHP* genes play important roles in floral organ determination (Pinyopich et al., 2003; Vrebalov et al., 2009; Chi et al., 2017). In *Arabidopsis*, a redundant roles of *SHP1/2* and *AG* were found to promote carpel development, and overexpression of *SHP2* was sufficient to rescue the stamen and carpel phenotypes in the *ag* mutant (stamens were replaced with petals and carpels were replaced by new abnormal flowers) (Bowman et al., 1989; Pinyopich et al., 2003). Transient knockdown of *NbSHP* in tobacco exhibited unfused pistils, and increased number of styles and stigmas (Fourquin and Ferrandiz, 2012). Ectopic expression of soybean *GmAGL1* in *Arabidopsis* resulted in petal-free flowers (Chi et al., 2017). Overexpression of grape *Vvmads1* (the ortholog of *SHP1/2*) in tobacco led to carpelloid sepals and stamenoid petals (Boss et al., 2001).

Cucumber (*Cucumis sativus* L.) is a world-wide cultivated vegetable crop bearing fleshy pepo fruit, which is developed from three-syncarpous inferior ovary (Che and Zhang, 2019). The fruit of North China cucumber can be generally divided into five developmental stages: immature green [0–9 days after anthesis days after anthesis (DAA)], breaker (12–15 DAA), turning (18–24 DAA), fully ripe (27–30 DAA), and senescence (35–43 DAA) (Leng et al., 2014). Cucumber fruits are consumed freshly or processed into pickles, and typically are harvested immature (about 7~14 DAA) (Weng et al., 2015). During summer production of cucumber, fruits left unpicked in time are prone to turn yellow and ripen on plants, which will greatly impair the commercial value and result in economic loss (Wang et al., 2013). Meanwhile, during cucumber post-harvest storage and transportation, unfavorable environmental conditions will cause immature fruits a series of senescence phenomena such as yellowing, accumulation of citrate, and tissue softening (Mainardi et al., 2006), which will adversely affect the market value. Therefore, identification of maturity-related genes is of great agricultural and economic importance for inhibiting fruit ripening on plants and delaying senescence of postharvest cucumbers.

As a non-climacteric fruit, it is abscisic acid (ABA), not ethylene, that was shown to promote fruit ripening on the cucumber plant (Nilsson, 2005; Hurr et al., 2009; Wang et al., 2013). To isolate putative fruit ripening related genes, we cloned a MADS-box gene *CsSHP* (the ortholog of *SHP*) in cucumber. *CsSHP* expression was highly enriched in reproductive organs of cucumber and were positively correlated with ABA accumulation during fruit maturation. Overexpression of *CsSHP* caused early flowering, precocious phenotypes, and ectopic capelloid organs in *Arabidopsis*. Biochemical analysis indicated that CsSHP directly interact with CsSEPs at protein level. Thus, CsSHP may participate in fruit maturation through the ABA pathway and floral organ specification *via* interaction with CsSEPs, which provides a possible target for genetic manipulation of fruit maturation progression to meet the different market demands in cucumber.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Cucumber (*Cucumis sativus* L.) inbred lines R1461 (Chinese long type) was used in this study, and grown in the experimental field of China Agricultural University at Beijing under standard greenhouse conditions. The *A. thaliana* Columbia (*Col*) ecotype and the double mutant *shp1 shp2*, and relevant transgenic plants were grown in soil at 22°C under 16 h/8 h light/dark condition in a growth chamber.

Sequence Alignment and Phylogenetic Analysis

A 711-bp fragment containing the complete *CsSHP* coding sequence was amplified from the young female buds using gene-specific primers (**Supplementary Table S2**). The gene structure of *CsSHP* was analyzed using online software GSDS 2.0 (http://gsds. cbi.pku.edu.cn/). Protein sequences of SHPs and other MADS-box genes from diverse plant species were obtained using the protein BLAST search (NCBI blast: https://blast.ncbi.nlm.nih.gov/Blast. cgi). Multiple sequence alignment was performed using CLUSTALW in MEGA5. The phylogenetic tree was generated using the neighbor-joining method in MEGA5 with 1,000 bootstrap replicates. The GenBank accession numbers for related proteins are listed in **Supplementary Table S1**.

Quantitative Real-Time Polymerase Chain Reaction

The young leaves, stems, tendrils, male flower buds, female flower buds, male flowers, female flowers, ovaries at anthesis, fruits at different developmental stages from cucumber, as well as the inflorescences of *Arabidopsis* were frozen in the liquid nitrogen and stored at −80°C until use. Total RNA was extracted with TRIzol reagent as described in the manufacturer's instructions (Waryoung, China, http://www. huayueyang.com/), and cDNA was synthesized using FastQuant RT Kit (Tiangen, China, http://www.tiangen.com/). Quantitative real-time PCR (qRT-PCR) was performed using TB Green™ Premix Ex Taq™ (Takara, Japan, http://www. takarabiomed.com.cn/) on the Applied Biosystems 7500 RT-qPCR system. *Ubiquitin extension protein* (*UBI* CsaV3_5G031430) and *ACTIN2* (AT3G18780) were used as the internal reference genes in cucumber and *Arabidopsis,* respectively (Wan et al., 2010). The relative expression was calculated according to the comparative cycle threshold (C_T) method (Schmittgen and Livak, 2008). Three biological replicates and three technical replicates were performed for each gene. The primer information is listed in **Supplementary Table S2**.

In Situ Hybridization

The 25-day-old shoot tips, male and female flower buds of R1461 were fixed with 3.7% formalin-acetic acid-alcohol (FAA) fixative. Sampling and recognition of flower buds at different developmental stages were performed as described (Bai et al., 2004). Sense and antisense probes were amplified with gene-specific primers containing SP6 and T7 RNA polymerase

binding sites, respectively. Sample fixation, sectioning and hybridization were performed as described previously (Zhang et al., 2013). The primer information is listed in **Supplementary Table S2**.

β-Glucuronidase Staining

β-glucuronidase (GUS) assay was performed according to the protocol as described previously (Liu et al., 2016). The *Arabidopsis* inflorescences and fruits at different developmental stage were fixed and incubated in GUS-staining solution at 37°C for 24–48 h until dyed blue. Stained samples were then cleared with 70% ethanol and observed by anatomic microscope (Leica DFC450, Germany). The experiment was repeated three times.

Subcellular Localization of *CsSHP*

The full-length CDS of *CsSHP* was fused into the pSUPER1300 vector to generate a CsSHP-GFP in-frame fusion protein. The empty pSUPER1300 vector was used as a positive control. *Agrobacterium tumefaciens* with the vectors were injected into the abaxial side of tobacco leaves (4–6 weeks old) by syringe as described previously (Schutze et al., 2009). After 48 h of infiltration, the fluorescence of the expressed GFP proteins was detected and captured under 488 nm excitation wavelength from Argon laser of fluorescence microscope (Leica sp5, Germany). The wavelength range for GFP is 495–545 nm.

Ectopic Expression in *Arabidopsis*

The binary vector pBI121 and pSUPER1300 were used for ectopic expression of *CsSHP* in *Arabidopsis Col* and *shp1 shp2*, respectively. The *CsSHP* CDS was cloned into pBI121 and pSUPER1300 through *Xba*I and *Sma*I cleavages sites, and *Hind*III and *Kpn*I cleavages sites to generate *CsSHP* overexpression constructs, respectively. *Col* and *shp1 shp2* plants were transformed by *A. tumefaciens* containing the pBI121 and pSUPER1300 recombinant construct through the floral dip method, respectively (Clough and Bent, 1998). Transgenic seeds were germinated on solid Murashige and Skoog (MS) medium with 50mg/L kanamycin and 25mg/L hygromycin, respectively. Resistant seedlings were transferred to soil and further verified by PCR and qRT-PCR. Three biological replicates and three technical replicates were performed for each gene. T2 transgenic plants were chosen for further phenotypic analysis and data statistics. The primers used for vector construction and transgene identification are given in **Supplementary Table S2**.

Measurement of Endogenous Hormones

About 0.3 g cucumber ovaries at different developmental stages were used as samples. The content of indole-3-acetic acid (IAA), zeatin riboside (ZR), abscisic acid (ABA) were measured using enzyme-linked immunosorbent assays according to methods previously described (Sun et al., 2016). Three biological replicates were performed.

Abscisic Acid and Auxin Treatment

Three groups, each consisting of 20 cucumber ovaries at 4 days before anthesis, were used as samples. The first group was sprayed with a solution of 200 mg/L ABA in 0.5% (v/v) Tween

20, the second group was sprayed with a solution of 50 mg/L IAA in 0.5% (v/v) Tween 20, and the third group was sprayed only with 0.5% (v/v) Tween 20 as control (Daminato et al., 2013). Three ovaries from each group were collected after 0, 1, 3, 12, and 24 h treatments. RNA of the collected ovaries was extracted, and the *CsSHP* expression was evaluated by qRT-PCR.

Yeast Two-Hybrid Assay

Full-length CDSs of *IND*, *SPT*, *CsSHP*, *CUM10*, *CsSEP2*, *CsSEP3*, and *CsSEP4* were cloned into pGADT7 (prey vector) or pGBKT7 (bait vector). All constructs were verified by sequencing and then transformed into yeast strain AH109. The yeast two-hybrid assays were conducted and protein interactions were analyzed on selective medium lacking Leu, Trp, His, and Ade (Ding et al., 2015). The primers for yeast two-hybrid assays are listed in **Supplementary Table S2**.

Firefly Luciferase Complementation Imaging Assay

CsSEP2, *CsSEP3*, *CsSEP4*, and *CUM10* full-length CDSs without stop codon were cloned in pCAMBIA1300-nLUC, and *CsSHP* full-length CDS with stop codon in pCAMBIA1300-cLUC. *A. tumefaciens* GV3101 strain carrying the above constructs was mixed in proportion and resuspended, then injected into tobacco (*N. benthamiana*) leaves by syringe. The interactions of the expressed fusion proteins were indicated by reconstituted LUC enzyme after 2–3 days of infiltration, and images were obtained using a chemiluminescent imaging system (Tanon 5200, China) as described (Chen et al., 2008). The primers for vector construction are given in **Supplementary Table S2**.

RESULTS

Identification of *CsSHP* in Cucumber

To isolate the *SHP* gene in cucumber, we performed a BLAST search in National Center for Biotechnology Information (NCBI) database, and found the cucumber protein (NP_001292697.1) displays the highest sequence homology (59.7% identity) to *Arabidopsis* SHP1. A reciprocal BLAST search was performed in The *Arabidopsis* Information Resource (TAIR) and cucurbit genomics database (CuGenDB), and the NP_001292697.1 (CsaV3_6G015770.1) was confirmed to be the SHP homolog in cucumber, therefore, we named it as *CsSHP* hereinafter. There were two *SHP* genes (*SHP1* and *SHP2*) in *Arabidopsis*, while only one *SHP* in cucumber. The genomic sequence of *CsSHP* is 8,591 bp, which is much longer than that of *SHP1* (4,058 bp) and *SHP2* (3,759 bp) in *Arabidopsis*. *CsSHP* is predicted to contain seven exons and six introns, with the first and second intron being particularly long (3,005 and 3,232 bp, respectively) (**Figure 1A**). The second intron of AG/PLE genes contains several conserved motifs and has been shown to be essential for fulfilling their proper functions (Causier et al., 2009). The extremely large first and second introns in *CsSHP* may imply the more complex functional regulation than its *Arabidopsis* counterpart.

The full-length coding sequence (CDS) of *CsSHP* was obtained from the female bud of cucumber inbred line R1461, which encodes a protein of 237 amino acids with a calculated molecular mass of 27.18 kD (**Supplementary S1**). A multiple sequence alignment of CsSHP and its homologs from other species indicated that these proteins contained the conserved MADS domain and K-box domain (**Figure 1B**). Phylogenetic analysis indicated that CsSHP is very close to ClSHP in watermelon, which was clustered with other known SHP proteins and located in the PLENA (PLE) lineages of AG subfamily, while CUM1 (the ortholog of AG in cucumber) was clustered with other AG proteins (**Figure 2**) (Kater et al., 1998).

Expression Analysis and Protein Localization of *CsSHP*

To explore the expression pattern of *CsSHP*, qRT-PCR was performed in different cucumber organs including young leaves, stems, tendrils, male buds, female buds, male flowers at anthesis, female flowers, and ovaries at anthesis. Transcripts of *CsSHP* were highly accumulated in reproductive organs such as male flowers, female flowers, and ovaries, but with low levels in vegetative organs including leaves, stems, and tendrils (**Figure 3A**). In the four floral organs at anthesis, *CsSHP* was specifically expressed in stamens of male flowers and stigma of female flowers, while very low levels in sepals and petals (**Figure 3B**).

A previous study showed that *SHP1/2* in *Arabidopsis* was specifically expressed in gynoecia, but not in stamens (Pinyopich et al., 2003; Colombo et al., 2010). To further detect the spatial and temporal expression pattern of *CsSHP*, *in situ* hybridization was applied to the shoot apex, ovaries, and flower buds at different developmental stages in cucumber. In the shoot apical meristem (SAM), floral meristem, and sepal primordia, *CsSHP* transcripts were undetectable (**Figures 3C–E**). *CsSHP* transcripts were first found in the initiating stamen primordia at stage 3 (**Figure 3F**), then *CsSHP* was expressed at stamen and carpel primordia from stage 4 to 6 (**Figures 3G, H**), and maintained its expression in developing stamens and degenerate carpels in male flowers (**Figures 3I, J**). In female flower buds, *CsSHP* was highly expressed in carpel primordia and decreasingly expressed in degenerate stamens (**Figures 3K–N**). Since stage 8, *CsSHP* transcripts were specifically detected at stigmas, placenta, and ovule primordia (**Figures 3O–Q**). No signal was detected upon hybridized with the sense probe control (**Figure 3R**). Transverse sections of ovaries showed significant enrichment of *CsSHP* signals at placenta, pseudoseptum, and ovules (**Figures 3S, T**).

To visualize the expression pattern at whole plant level, transgenic *Arabidopsis* lines expressing β-glucuronidase (GUS) driven by the 1,716 bp *CsSHP* promoter fragment were generated. Unlike the specific expression of *SHP2* throughout the developing gynoecium in *Arabidopsis* (Colombo et al., 2010), GUS signal of *CsSHP* was found strongly in stamens as well as the valve margin of siliques, but not in developing gynoecium (**Figures 3U–Z**).

To explore the subcellular localization, CsSHP was fused with GFP under the control of a pSUPER promoter, and transiently expressed in tobacco leaves. Confocal green fluorescence imaging
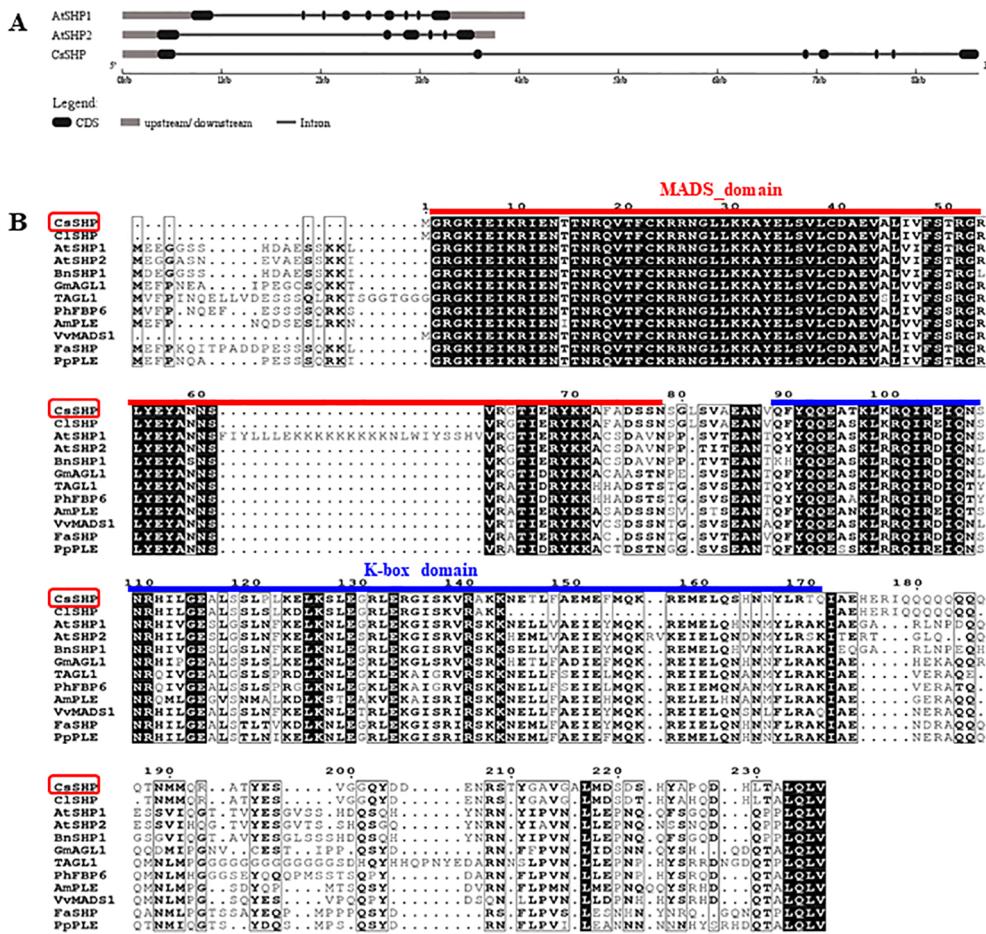
FIGURE 1 | Gene structure and sequence alignment. **(A)** Gene structure analysis of *SHP1*, *SHP2*, and *CsSHP*. **(B)** Multiple sequence alignment of CsSHP and orthologs from related plant species. The red and blue lines indicate the conserved MADS domain and K-box domain, respectively.

revealed that CsSHP-GFP fusion protein was located to the nucleus, whereas free GFP was distributed throughout the cell (**Figures 3A, B**).

Ectopic Expression of *CsSHP* Led to Early Flowering and Disturbed Floral Organ Development in *Arabidopsis*

In *Arabidopsis*, the mature siliques in the double mutant *shp1 shp2* plant were unable to dehisce due to failure of dehiscence zone formation (Liljegren et al., 2000). To explore the function of *CsSHP*, we first transformed *CsSHP* driven by the pSUPER promoter into *shp1 shp2* mutant plants. A total of 10 transgenic plants were obtained and the T2 plants were used for further characterization. Our data showed that ectopic expression of *CsSHP* was unable to rescue the indehiscence phenotype of *shp1 shp2* mutant plant, but instead resulted in early flowering in *Arabidopsis* (**Supplementary Figures S1A, B**). Next, we transformed the *CsSHP* driven by the cauliflower mosaic virus 35S (CaMV 35S) into wild-type *Arabidopsis* (*35S:: CsSHP/Col*). A total of 29 transgenic lines were generated. Based on *CsSHP* expression levels, three representative lines (#41, #45,

#53) (T2) were chosen for further phenotypic analysis and data statistics. Compared to the wild-type control (*Col*), the expression of *CsSHP* was dramatically increased in the transgenic lines (**Supplementary Figure S2A**). Overexpression of *CsSHP* resulted in early flowering (**Figure 4A**). Quantification analysis indicated that the days to bolting was 27.2 ± 2.1 in *Col*, while that of overexpression lines varied from 21.5 ± 1.3 to 23.3 ± 1.0 (**Figure 4B**). Similarly, the number of rosette leaves upon bolting, as well as the days to the 1st flower opening were significantly reduced in the transgenic lines (**Supplementary Figures S2B, C**). Moreover, ectopic expression of *CsSHP* accelerated the progression of reproductive growth in *Arabidopsis*. Under the same conditions, wild-type siliques in the main inflorescence were still green, while many of those in the transgenic plants had turned yellow and or even cracked (**Figure 4C**). Data statistics showed that the days to 1st silique yellowing and the days from anthesis to silique cracking were significantly shorter in transgenic lines than *Col* plants (**Figures 4D, E**).

Much more dramatic changes were found in flower organ development in the *35S::CsSHP* transgenic lines. As compared to
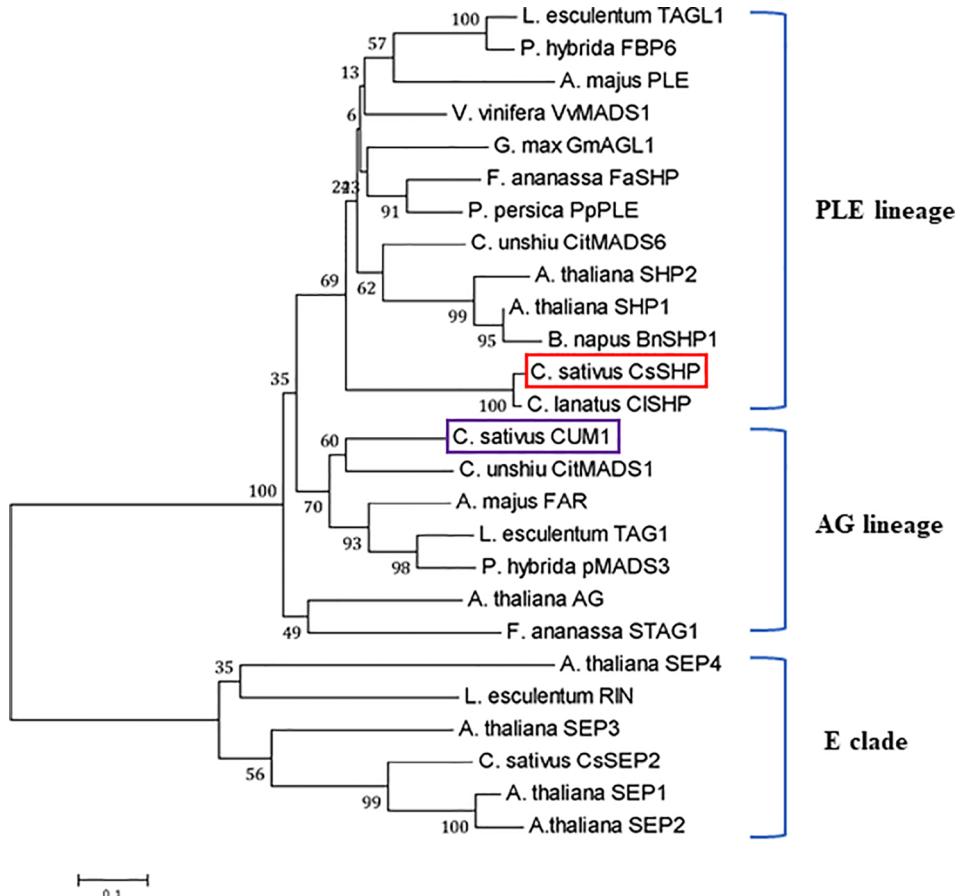
FIGURE 2 | Phylogenetic analysis of CsSHP and its homologous proteins. The scale bar represents 0.1 substitutions per site. The red and purple squares indicate CsSHP and CUM1, respectively. E lineage proteins were used as outgroups.

the wild-type control, some flower buds in transgenic plants were precociously opened (**Figures 4F, G**), flower patterning was disrupted, and the inflorescence meristem prematurely terminated (**Figures 4H, I**). For each individual flower, the flower patterning of sepal, petal, stamen, carpel was disturbed upon *CsSHP* overexpression. Some flowers in transgenic lines lack petals and stamens, consisting of only gynoecia and sepals, in which the gynoecia formed a longitudinal cleft to expose the placenta and ovules, and sepals were carpelloid (**Figure 4J**). Sometimes, one sepal of the flower was replaced by an ectopic gynoecium containing placenta tissue and ovules (**Figures 4K, L**). In more severe cases, a misshapen flower with carpelloid sepals grew from the base of another abnormal silique (**Figure 4M**), or an inflorescence grow out from the base of an ectopic carpel bearing ovules, and almost all flowers on the inflorescences were abnormal (**Figure 4N**). Scanning electron microscopy images showed that wild-type sepals were long and narrow, while that of the transgenic plants were short and round, with ectopic placenta and ovules in the inner side of the sepal (**Figures 4O, P**). Abnormal flowers with only gynoecium and carpelloid sepal grew from the inner base of another gynoecium or another flower (**Figures 4Q–S**). Statistic data showed that

average of 51.6% plants displayed abnormal flowers in the transgenic lines, in which 8% of flowers were defected, as compared to none in the wild-type control (**Supplementary Figure S2D**).

Based on the carpelloid organ phenotype, a set of carpel related genes including *AG*, *SEP1/2/3/4*, and *STK* were chosen for expression analysis (Pelaz et al., 2000; Favaro et al., 2003). As compared to the wild-type control, the expression of *AG* was significantly elevated, whereas that of *SEP2*, *SEP3*, *SEP4*, and *STK* was dramatically decreased in the *CsSHP* transgenic plants (**Figure 4T**).

In addition, unlike the stretching and flat wild-type leaves, there was a considerable proportion of crooked rosette and cauline leaves curling upwards and inward in transgenic lines (**Supplementary Figure S3**).

Interactions Of CsSHP With CsSEPs

SEP1/2/3/4 are a class of organ-identity genes that are required for development of sepals, petals, stamens, and carpels in *Arabidopsis* (Pelaz et al., 2000), and SEP proteins are thought to act as a "bridge" allowing the formation of higher order complex with the floral organ identity MADS-box factors
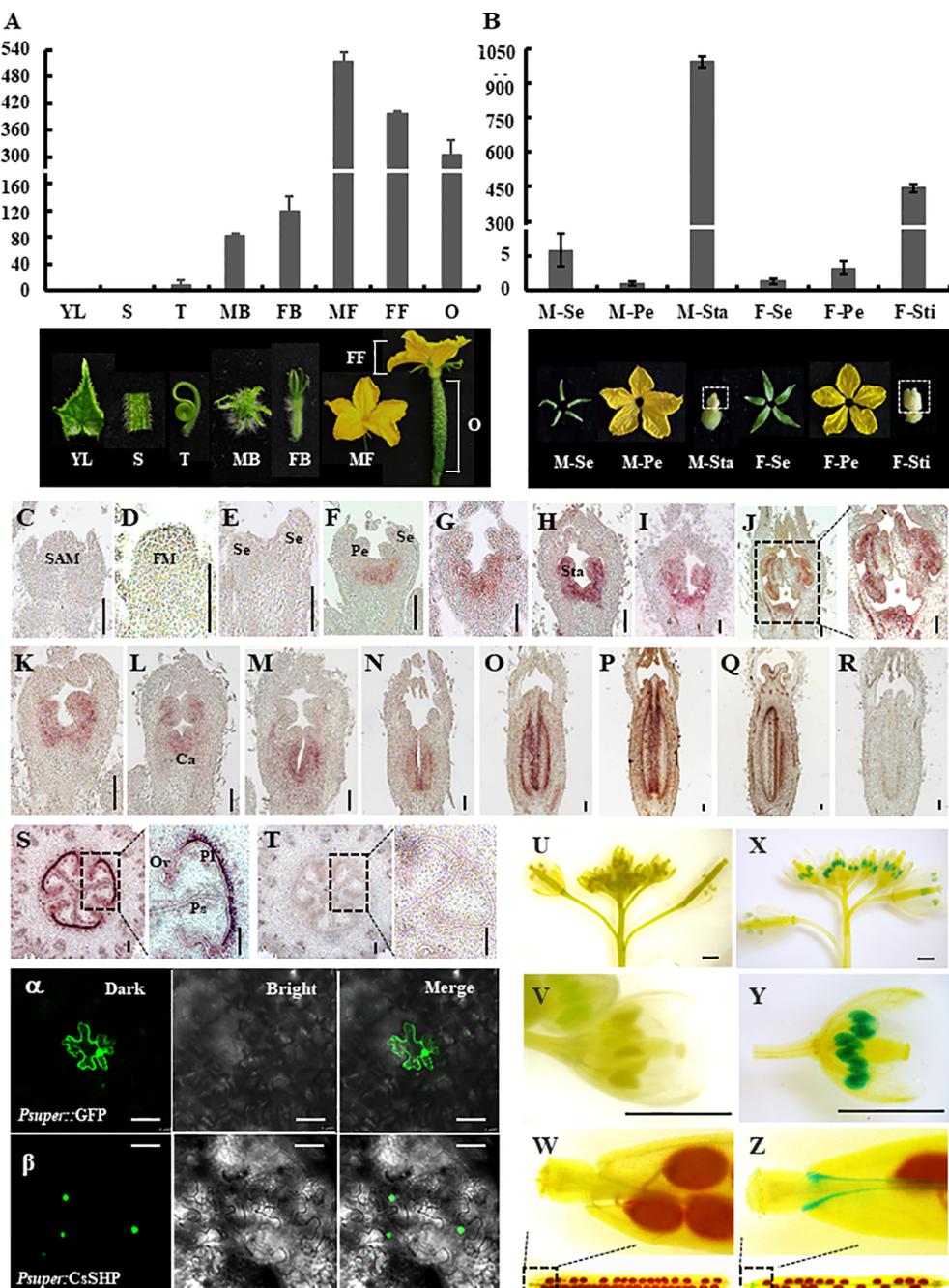
FIGURE 3 | Expression analysis and subcellular localization of *CsSHP*. **(A)** quantitative real-time (qRT)-PCR analysis of *CsSHP* expression in different cucumber organs. YL, young leaves; S, stems; T, tendrils; MB, male buds; FB, female buds; MF, male flowers at anthesis; FF, female flowers at anthesis; O, ovaries at anthesis. Three biological replicates and three technical replicates were performed. **(B)** qRT-PCR analysis of *CsSHP* expression in different floral organs at anthesis. M-Se, sepals from male flowers; M-Pe, petals from male flowers; M-Sta, stamens from male flowers; F-Se, sepals from female flowers; F-Pe, petals from female flowers; F-Sti, stigmas from female flowers. Three biological replicates and three technical replicates were performed. Pictures of the corresponding organs are displayed at the bottom. **(C–T)** *In situ* hybridization of *CsSHP* in young organs of cucumber. Scale bar = 100 μm. **(C)** Shoot apical meristem (SAM). **(D–H)** Longitudinal sections of male flower buds at stages 1–5 in inbred line R1461. **(I, J)** Male flower buds at stages 8 and 9 in R1461 line. **(K–Q)** Longitudinal sections of female flower buds at different developmental stages in inbred line R1461. **(S)** Transection of ovary 10 days before anthesis. **(R, T)** The sense *CsSHP* probe was hybridized as a negative control. FM, floral meristem; Se, sepal; Pe, petal; Sta, stamen; Ca, carpel; Pl, placenta; Ps, pseudoseptum; Ov, ovule. **(U–Z)** β-Glucuronidase (GUS) signals of *CsSHP* in *Arabidopsis*. Scale bar = 1 mm. **(U–W)** Negative control plants showed no GUS signal. **(X, Y)** GUS signals were highly enriched in stamens. **(Z)** GUS signals were detected in valve margin of silique. The experiment was repeated three times. **(α, β)** Subcellular localization of CsSHP. GFP signals were localized in the nucleus. GFP driven by the pSUPER promoter was used as a positive control. GFP is shown in green. The left, middle, and right panels represent pictures taken under dark field, bright field, and merge views, respectively. Scale bar = 50 μm.
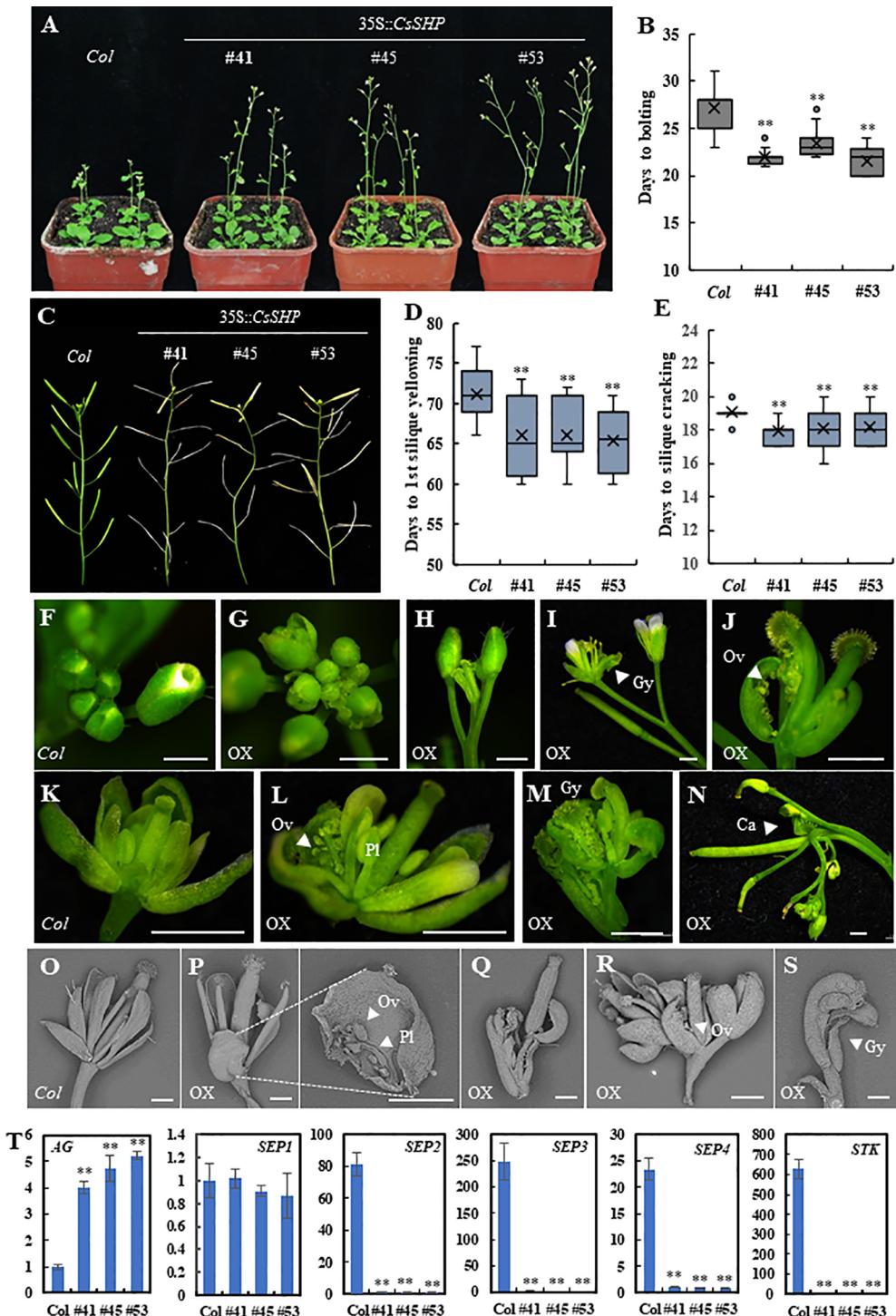
FIGURE 4 | Phenotypic characterization of *35S::CsSHP* transgenic plants in *Arabidopsis*. **(A)** Representative images of *35S::CsSHP* transgenic plants indicated early flowering in *Arabidopsis*. **(B)** Box-plot of the days to bolting in *Col* and *35S::CsSHP* transgenic lines. **(C)** Overexpression of *CsSHP* led to precocious phenotype in *Arabidopsis*. **(D, E)** Quantification of the days to yellowing of the 1st silique **(D)** and the days from anthesis to silique cracking **(E)**. The data were the average of 20 plants for each line. **(F–S)** Overexpression of *CsSHP* caused defected floral organ development in *Arabidopsis*. **(F–N)** Anatomic micrographs of flowers and inflorescences. Scale bar = 1 mm. **(O–S)** Scanning electron micrographs of flowers and siliques. Scale bar = 500 μm. **(T)** The expression levels of *AG*, *SEP1*, *SEP2*, *SEP3*, *SEP4,* and *STK* in *35S::CsSHP Arabidopsis* flowers. Gy, gynoecium; Ov, ovule; Pl, placenta; Ca, carpel **, t-test (p < 0.01).

(Favaro et al., 2003). *In vitro* and *in vivo* evidence was provided for the existence of SEP, STK, and/or AG and SHP protein complexes in promoting the *Arabidopsis* ovule identity (Mendes et al., 2013). In cucumber, constitutive expression of *CUM1* (AG ortholog) resulted in sepals transformed into carpelloid structures, and petals reduced significantly in size or completely absent (Kater et al., 2001). CUM10 (STK ortholog) mediates floral organ identity in cucumber, and ectopic expression of *CUM10* resulted in partial transformation of petals into antheroid structures in petunia (Kater et al., 1998). *CsSEP2* was shown to participate in floral organ development in cucumber, since abolishment of the transcriptional activity of *CsSEP2* led to increased floral organ size with disturbed floral patterning (Wang et al., 2016). To explore the existence of possible interactions in cucumber, a yeast two-hybrid assay was performed between CsSHP and cucumber homologs of AG, SEPs, and STK. Our data showed that CsSHP displayed strong interactions with CsSEP2, CsSEP3, CsSEP4, and a weak interaction with CUM10, while CsSHP could not interact with CUM1, neither with itself to form homodimer (**Figure 5A**). To verify the protein interactions *in vivo*, a LUC complementation imaging assay was performed in the abaxial leaf epidermis of tobacco (**Figure 5B**). The reconstituted LUC enzyme can be detected by luminometer from combinations of CsSHP with CsSEP2, CsSEP3, or CsSEP4, but not from combination of CsSHP with CUM1 or CUM10, indicating that CsSHP interacts with CsSEPs to form multimeric protein complex in cucumber.

CsSHP Expression Is Correlated With Fruit Maturation in Cucumber

In climacteric tomato, *TAGL1* was highly expressed in flowers at anthesis and in fruits at red ripe stage (Gimenez et al., 2010). Similarly, during the development of non-climacteric citrus fruits, the transcript levels of *CsMADS6* increased gradually, peaking at 180 DAA (breaker stage) and declining at 220 DAA (Lu et al., 2018). To understand whether *CsSHP* expression is correlated with fruit maturation in cucumber, qRT-PCR analysis was performed in cucumber fruit at fourteen developmental points. As shown in **Figure 6A**, the *CsSHP* expression showed a slight downward trend from 0.3 cm flower buds (14 days before anthesis) to commercially mature fruits (nine DAA). From 9 to 15 DAA, *CsSHP* transcripts increased threefold during the 6 days, and the highest accumulation of transcripts was detected at 21 DAA when the pericarp began to yellow. After 21 DAA, *CsSHP* expression gradually declined as fruit maturation progression and the pericarp became more and more yellow.

Cucumber is classified as a non-climacteric fruit, and endogenous ethylene displayed no influence on the postharvest yellowing of cucumber (Nilsson, 2005). However, abscisic acid (ABA) was shown to play important and direct roles in regulating cucumber fruit development and ripening. Exogenous ABA application at the turning stage promoted fruit ripening in cucumber (Wang et al., 2013). Therefore, the content of ABA, as well as the indole-3-acetic acid (IAA) and zeatin riboside (ZR), was measured in fruits of the variety R1461 from 9 to 30 DAA, when *CsSHP* expression changed drastically. Our data showed that the content of ABA increased continuously from 9 to 30 DAA as fruit ripening progression, with a dramatic upregulation from 15 to 30 DAA, while that of IAA and ZR displayed a mild increase from 9 to 15 DAA, and then decreased (**Figure 6B**), confirming that ABA may play key roles in mediating fruit ripening in cucumber. Given that the *CsSHP* expression and ABA level showed similar trend as fruit maturation, we next examined *CsSHP* response to ABA treatment in ovaries of 4 days before anthesis. Our data indicated that *CsSHP* transcription was significantly induced after ABA treatment for 3 hours, and remained at 2–3 times upregulation to 24 h (**Figure 6C**), suggesting that CsSHP may be involved in fruit maturation in cucumber through the ABA pathway.

DISCUSSION

MADS-box genes are critical transcription factors, participating in virtually all aspects of plant development, especially during flower and fruit development (Smaczniak et al., 2012; Wang et al., 2015). Previous study revealed the expression pattern of 43 MADS-box genes in cucumber (Hu and Liu, 2012). *CUM1*, *CUM10*, *CUM26*, *CsAP3*, and *CsSEP2* are currently known cucumber MADS-box genes that play essential roles in flower and fruit development (Kater et al., 1998; Kater et al., 2001; Sun et al., 2016; Wang et al., 2016), whereas there are still a considerable portion of MADS-box genes need to be functionally characterized. In this study, we identified the *CsSHP* gene in cucumber and revealed the roles in fruit maturation and floral organ determination.

CsSHP is a member of MADS-box family and has the conserved MADS domain and K-box domain. The MADS domain is responsible for binding downstream target DNA, and the K-box domain acts on protein-protein interactions (Shore and Sharrocks, 1995; Yang et al., 2003). The CDS length of *CsSHP* and *SHP1/2* is almost equivalent, while the genomic length of *CsSHP* gene is about twice that of *SHP1/2*, due to extremely long first and second introns (**Figure 1A**). Expression analyses by RT-PCR and *in situ* hybridization showed that *CsSHP* was specifically expressed in reproductive tissues including stamens of male flowers, stigma of female flowers, and ovaries (**Figure 3**). Such expression is similar to that of *TAGL1* in tomato and *NbSHP* in tobacco, but different from *SHP2* in *Arabidopsis*, which is specifically expressed in gynoecium and absent in stamen (Vrebalov et al., 2009; Gimenez et al., 2010; Fourquin and Ferrandiz, 2012). Interestingly, GUS signals driven by *CsSHP* promoter were detected only in stamens and valve margin in *Arabidopsis*, but not in developing gynoecium (**Figure 3**). There are two possible reasons for such discrepancy. One is the different flower structure in *Arabidopsis* (complete flower) and cucumber (unisexual flower), and the other is that the promoter sequence used for driving the GUS signal was unable to confer the correct expression pattern of *CsSHP*. Considering that the second intron of AG/PLE genes contain multiple regulatory elements that determine the proper spatial and temporal expression (Liu and Liu, 2008; Gu et al., 2018), it is more plausible that the expression pattern of *CsSHP* is coordinately controlled by its promoter and introns.
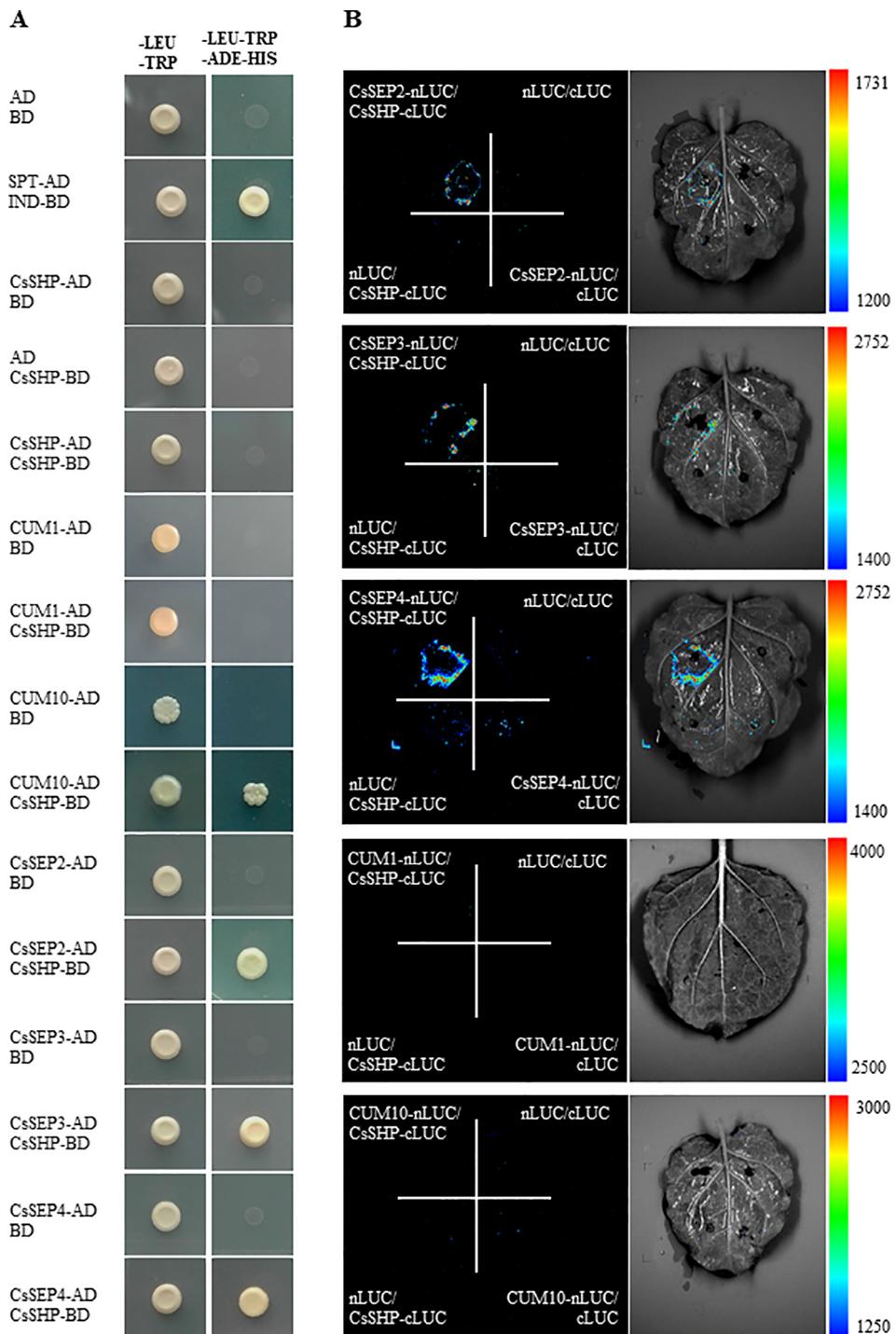
FIGURE 5 | Physical interactions of CsSHP with CsSEP2, CsSEP3, and CsSEP4. **(A)** Yeast two-hybrid assay. The combination of SPT-AD and IND-BD was used as a positive control. **(B)** Luciferase (LUC) complementation imaging analysis.

Overexpression of *CsSHP* in wild-type *Arabidopsis* resulted on ectopic carpelloid organs with the characteristics of gynoecium: stigma-like tissue at the top, placenta tissue in the middle, and infertile ovules along the valve margin (**Figure 4**).

The inflorescence or flower patterning were often terminated in an ectopic carpel, showing a phenotype of "flower in carpel" or "inflorescence in carpel" (**Figure 4**). Although some phenotypes like early flowering, curly leaves and prematurely open flower
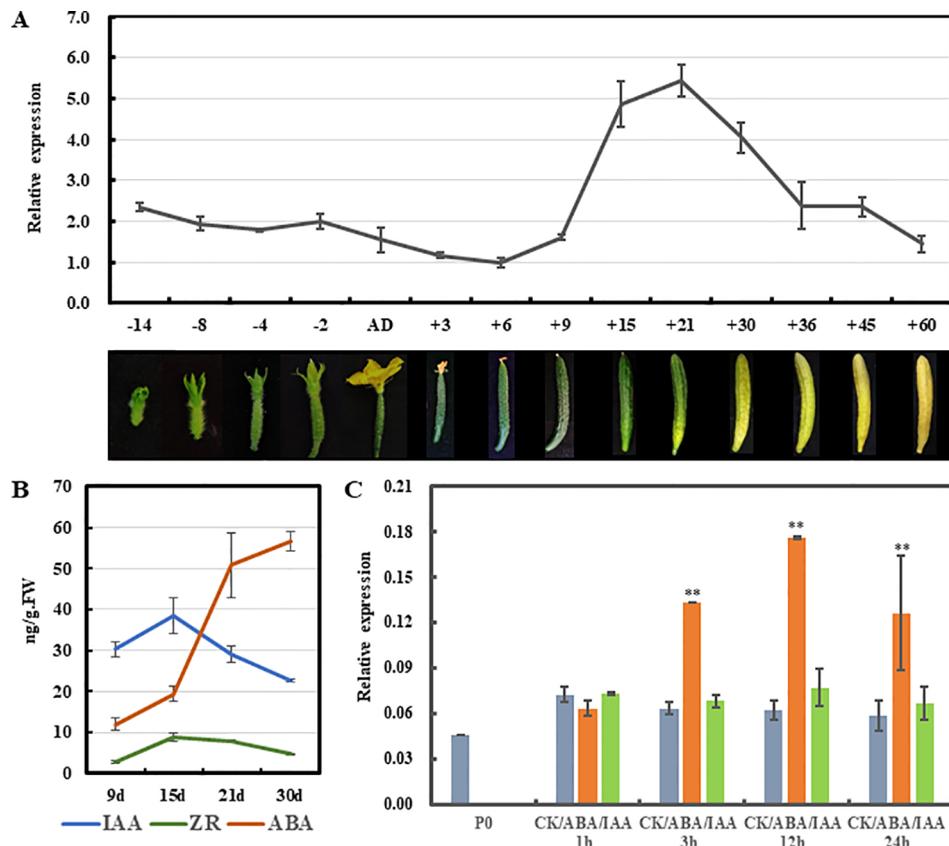
FIGURE 6 | Expression analysis of *CsSHP* during fruit maturation in cucumber. **(A)** *CsSHP* expression in fruits at different developmental stages in inbred line R1461. Fruit pictures of the corresponding developmental stages are displayed at the bottom in different proportions. **(B)** Measurements of three endogenous hormones in fruit from 9 to 30 days after anthesis (DAA). Three biological replicates were performed. **(C)** Expression response of *CsSHP* to abscisic acid (ABA) or indole-3-acetic acid (IAA) treatment in cucumber ovaries of 4 days before anthesis. Three biological replicates and three technical replicates were performed **, t-test ($p < 0.01$).

buds, are similar to those previously described upon overexpression of *SHP* orthologs, such as *Arabidopsis SHP1/2* and tomato *TAGL1,* in *Arabidopsis* (Favaro et al., 2003; Vrebalov et al., 2009), the carpelloid phenotype caused by *CsSHP* overexpression is more severe. For example, *35S::SHP1/2* in *Arabidopsis* showed a transformation of petals toward stamens and a partial conversion of sepals toward carpels, with stigmatic papillae on it (Favaro et al., 2003). However, in the *Arabidopsis* inflorescences of *35S::CsSHP* plants, ectopic carpels may grow anywhere of the flower (**Figure 4**). The carpelloid phenotype in *35S::CsSHP* transgenic plants suggested the function of *CsSHP* in specifying carpel identity in cucumber. Consistently, the expression of *AG* was significantly elevated, while that of STK was almost abolished in the transgenic flowers (**Figure 4**), which may due to the ectopic carpelloid organs with infertile ovules. Interestingly, the expression of *SEP2*, *SEP3* and *SEP4* was dramatically decreased in the *35S::CsSHP* transgenic flowers (**Figure 4**). However, without a time course analysis in *CsSHP* inducible lines, we were unable to differentiate the expression change of *CsSEP2/3/4* is a cause or a result of floral patterning defect. Biochemical data showed that CsSHP interacts with CsSEP2/3/4 at protein level (**Figure 5**). Loss of *CsSEP2*

function led to disturbed flower patterning with enlarged floral organs in cucumber (Wang et al., 2016). Therefore, we hypothesized that CsSHP may specify carpel identity in cucumber through interacting with CsSEP2 to form protein complex. In fact, SHP homologs were shown to interact with SEP-like proteins in *Arabidopsis*, tomato, and soybean (Favaro et al., 2003; Leseberg et al., 2008; Mendes et al., 2013; Chi et al., 2017), indicating the interactions between SHP and SEPs are relatively conserved among species.

Cucumber fruit is harvested immature and consumed freshly or as processed pickles. Fruit ripening is important for seed maturation, but has adverse effect on cucumber production and post-harvest shelf life. Unlike the dry pod in *Arabidopsis*, cucumber fruit has no cracking characteristics, and fruit ripening involved in changes of color, texture, and aroma (Seymour et al., 2008). As a non-climacteric fruit, ethylene has no influence on the postharvest yellowing of cucumber (Nilsson, 2005), but instead, ABA application at the turning stage promoted fruit ripening (Wang et al., 2013). Here, we found that ectopic expression of *CsSHP* resulted in early flowering and accelerated ripening in *Arabidopsis* (**Figure 4**). Moreover, *CsSHP* expression is significantly increased during

the fruit yellowing from 9 to 21 DAA, concomitant with the dramatic increase of ABA level (**Figure 6**), implying the positive roles of *CsSHP* in cucumber ripening. Further, *CsSHP* expression was found to be induced upon exogenous ABA application (**Figure 6D**). Hence, it is plausible to speculate that high level of ABA from commodity fruit stage (nine DAA) induces *CsSHP* expression, which positively modulates carotenoid metabolism and ripening-related genes, just like AGL1 in tomato and CsMADS6 in citrus (Vrebalov et al., 2009; Lu et al., 2018), to promote fruit yellowing and fruit ripening in cucumber. Further studies are needed to test above hypothesis using transgenic cucumbers through CRISPR/Cas9 techniques (Hu et al., 2017).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ZC and XZ conceived this project. ZC, SZ, XL, GC, ZW, RG, JS, WS performed the experiments, ZC and XZ wrote the manuscript, XL, ZZ and DH contributed to critical discussions. All authors read and approved the final version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01781/full#supplementary-material

REFERENCES

Bai, S. L., Peng, Y. B., Cui, J. X., Gu, H. T., Xu, L. Y., Li, Y. Q., et al. (2004). Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus L.*). *Planta* 220, 230–240. doi: 10.1007/s00425-004-1342-2

Boss, P. K., Vivier, M., Matsumoto, S., Dry, I. B., and Thomas, M. R. (2001). A cDNA from grapevine (Vitis vinifera L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. *Plant Mol. Biol.* 45, 541–553. doi: 10.1023/A:1010634132156

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1989). Genes directing flower development in *Arabidopsis*. *Plant Cell.* 1, 37–52. doi: 10.1105/tpc.19.00276

Bowman, J. L., Drews, G. N., and Meyerowitz, E. M. (1991). Expression of the *Arabidopsis* floral homeotic gene agamous is restricted to specific cell-types late in flower development. *Plant Cell.* 3, 749–758. doi: 10.1105/tpc.3.8.749

Causier, B., Bradley, D., Cook, H., and Davies, B. (2009). Conserved intragenic elements were critical for the evolution of the floral C-function. *Plant J.* 58, 41–52. doi: 10.1111/j.1365-313X.2008.03759.x

Che, G., and Zhang, X. (2019). Molecular basis of cucumber fruit domestication. *Curr. Opin. Plant Biol.* 47, 38–46. doi: 10.1016/j.pbi.2018.08.006

Chen, H. M., Zou, Y., Shang, Y. L., Lin, H. Q., Wang, Y. J., Cai, R., et al. (2008). Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol.* 146, 368–376. doi: 10.1104/pp.107.111740

Chi, Y. J., Wang, T. T., Xu, G. L., Yang, H., Zeng, X. R., Shen, Y. X., et al. (2017). GmAGL1, a MADS-Box gene from soybean, is involved in floral organ identity and fruit dehiscence. *Front. Plant Sci.* 8, 175. doi: 10.3389/fpls.2017.00175

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis* thaliana. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls -genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0

Colombo, M., Brambilla, V., Marcheselli, R., Caporali, E., Kater, M. M., and Colombo, L. (2010). A new role for the SHATTERPROOF genes during *Arabidopsis* gynoecium development. *Dev. Biol.* 337, 294–302. doi: 10.1016/j.ydbio.2009.10.043

Daminato, M., Guzzo, F., and Casadoro, G. (2013). A SHATTERPROOF-like gene controls ripening in non-climacteric strawberries, and auxin and abscisic acid antagonistically affect its expression. *J. Exp. Bot.* 64, 3775–3786. doi: 10.1093/jxb/ert214

Ding, L., Yan, S. S., Jiang, L., Zhao, W. S., Ning, K., Zhao, J. Y., et al. (2015). HANABA TARANU (HAN) bridges meristem and organ primordia boundaries through PINHEAD, JAGGED, BLADE-ON-PETIOLE2 and CYTOKININ OXIDASE 3 during flower development in *Arabidopsis*. *Plos Genet.* 11, e1005479. doi: 10.1371/journal.pgen.1005479

Dreni, L., and Kater, M. M. (2014). MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol.* 201, 717–732. doi: 10.1111/nph.12555

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., et al. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell.* 15, 2603–2611. doi: 10.1105/tpc.015123

Ferrandiz, C., and Fourquin, C. (2014). Role of the FUL-SHP network in the evolution of fruit morphology and function. *J. Exp. Bot.* 65, 4505–4513. doi: 10.1093/jxb/ert479

Fourquin, C., and Ferrandiz, C. (2012). Functional analyses of AGAMOUS family members in *Nicotiana benthamiana* clarify the evolution of early and late roles of C-function genes in eudicots. *Plant J.* 71, 990–1001. doi: 10.1111/j.1365-313X.2012.05046.x

Fourquin, C., del Cerro, C., Victoria, F. C., Vialette-Guiraud, A., de Oliveira, A. C., and Ferrandiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* Genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.217570

Gimenez, E., Pineda, B., Capel, J., Anton, M. T., Atares, A., Perez-Martin, F., et al. (2010). Functional Analysis of the arlequin mutant corroborates the essential role of the ARLEQUIN/TAGL1 gene during reproductive development of tomato. *Plos One* 5, e14427. doi: 10.1371/journal.pone.0014427

Giovannoni, J. J. (2007). Fruit ripening mutants yield insights into ripening control. *Curr. Opin. Plant Biol.* 10, 283–289. doi: 10.1016/j.pbi.2007.04.008

Gu, R., Liu, X. F., Zhao, W. S., Yan, S. S., Sun, L. H., Wu, B. N., et al. (2018). Functional characterization of the promoter and second intron of CUM1 during flower development in Cucumber (*Cucumis sativus L.*). *Hortic. Plant J.* 4, 103–110. doi: 10.1016/j.hpj.2018.03.004

Hu, L. F., and Liu, S. Q. (2012). Genome-wide analysis of the MADS-box gene family in cucumber. *Genome* 55, 245–256. doi: 10.1139/g2012-009

Hu, B. W., Li, D. W., Liu, X., Qi, J. J., Gao, D. L., Zhao, S. Q., et al. (2017). Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 System. *Mol. Plant.* 10, 1575–1578. doi: 10.1016/j.molp.2017.09.005

Hurr, B. M., Huber, D. J., Vallejos, C. E., and Talcott, S. T. (2009). Developmentally dependent responses of detached cucumber (*Cucumis sativus L.*) fruit to exogenous ethylene. *Postharv. Biol. Tec.* 52, 207–215. doi: 10.1016/j.postharvbio.2008.12.006

Itkin, M., Seybold, H., Breitel, D., Rogachev, I., Meir, S., and Aharoni, A. (2009). TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. *Plant J.* 60, 1081–1095. doi: 10.1111/j.1365-313X.2009.04064.x

Jofuku, K. D., Denboer, B. G. W., Vanmontagu, M., and Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene Apetala2. *Plant Cell* 6, 1211–1225. doi: 10.1105/tpc.6.9.1211

Kater, M. M., Colombo, L., Franken, J., Busscher, M., Masiero, S., Campagne, M. M. V., et al. (1998). Multiple AGAMOUS homologs from cucumber and petunia differ in their ability to induce reproductive organ fate. *Plant Cell* 10, 171–182. doi: 10.1105/tpc.10.2.171

Kater, M. M., Franken, J., Carney, K. J., Colombo, L., and Angenent, G. C. (2001). Sex determination in the monoecious species cucumber is confined to specific floral whorls. *Plant Cell* 13, 481–493. doi: 10.1105/tpc.13.3.481

Leng, P., Yuan, B., and Guo, Y. D. (2014). The role of abscisic acid in fruit ripening and responses to abiotic stress. *J. Exp. Bot.* 65, 4577–4588. doi: 10.1093/jxb/eru204

Leseberg, C. H., Eissler, C. L., Wang, X., Johns, M. A., Duvall, M. R., and Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. *J. Exp. Bot.* 59, 2253–2265. doi: 10.1093/jxb/ern094

Liljegren, S. J., Ditta, G. S., Eshed, H. Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089

Liu, Z. R., and Liu, Z. C. (2008). The second intron of AGAMOUS drives carpel- and stamen-specific expression sufficient to induce complete sterility in *Arabidopsis*. *Plant Cell Rep.* 27, 855–863. doi: 10.1007/s00299-008-0511-3

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

Lu, S. W., Zhang, Y., Zhu, K. J., Yang, W., Ye, J. L., Chai, L. J., 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

Mainardi, J. A., Purgatto, E., Vieira, A.Jr., Bastos, W. A., Cordenunsi, B. R., Oliveira do Nascimento, J. R., et al. (2006). Effects of ethylene and 1-methylcyclopropene (1-MCP) on gene expression and activity profile of alpha-1,4-glucan-phosphorylase during banana ripening. *J. Agric. Food Chem.* 54, 7294–7299. doi: 10.1021/jf061180k

Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., et al. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *Plant Cell* 25, 2560–2572. doi: 10.1105/tpc.112.108688

Mizukami, Y., and Ma, H. (1992). Ectopic expression of the floral homeotic gene agamous in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* 71, 119–131. doi: 10.1016/0092-8674(92)90271-d

Mizukami, Y., and Ma, H. (1995). Separation of ag function in floral meristem determinacy from that in reproductive organ identity by expressing antisense ag RNA. *Plant Mol. Biol.* 28, 767–784. doi: 10.1007/BF00042064

Nilsson, T. (2005). Effects of ethylene and 1-MCP on ripening and senescence of European seedless cucumbers. *Postharv. Biol. Tec.* 36, 113–125. doi: 10.1016/j.postharvbio.2004.11.008

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88. doi: 10.1038/nature01741

Robles, P., and Pelaz, S. (2005). Flower and fruit development in *Arabidopsis thaliana*. *Int. J. Dev. Biol.* 49, 633–643. doi: 10.1387/ijdb.052020pr

Roeder, A. H., and Yanofsky, M. F. (2006). Fruit development in *Arabidopsis*, *Arabidopsis Book*. vol. 4, e0075. doi: 10.1199/tab.0075

Ruelens, P., Zhang, Z. C., van Mourik, H., Maere, S., Kaufmann, K., and Geuten, K. (2017). The origin of floral organ identity quartets. *Plant Cell* 29, 229–242. doi: 10.1105/tpc.16.00366

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Schutze, K., Harter, K., and Chaban, C. (2009). Bimolecular fluorescence complementation (BiFC) to study protein-protein interactions in living plant cells. *Methods Mol Biol.* 479, 189–202. doi: 10.1007/978-1-59745-289-2_12

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

Shore, P., and Sharrocks, A. D. (1995). The MADs-Box family of transcription factors. *Eur. J. Biochem.* 229, 1–13. doi: 10.1111/j.1432-1033.1995.tb20430.x

Smacszniak, C., Immink, R. G. H., Angenent, G. C., and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139, 3081–3098. doi: 10.1242/dev.074674

Sun, J. J., Li, F., Wang, D. H., Liu, X. F., Li, X., Liu, N., et al. (2016). CsAP3: a cucumber homolog to *Arabidopsis* APETALA3 with novel characteristics. *Front. Plant Sci.* 7, 1181. doi: 10.3389/fpls.2016.01181

Theissen, G., and Saedler, H. (2001). Plant biology - floral quartets. *Nature* 409, 469–471. doi: 10.1038/35054172

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z. J., et al. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Front. Plant Sci.* 5, 227. doi: 10.3389/fpls.2014.00227

Vrebalov, J., Pan, I. L., Arroyo, A. J. M., McQuinn, R., Chung, M., Poole, M., et al. (2009). Fleshy fruit expansion and ripening are regulated by the tomato SHATTERPROOF gene TAGL1. *Plant Cell* 21, 3041–3062. doi: 10.1105/tpc.109.066936

Wan, H. J., Zhao, Z. G., Qian, C. T., Sui, Y. H., Malik, A. A., and Chen, J. F. (2010). Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. *Anal. Biochem.* 399, 257–261. doi: 10.1016/j.ab.2009.12.008

Wang, Y. P., Wang, Y., Ji, K., Dai, S. J., Hu, Y., Sun, L., et al. (2013). The role of abscisic acid in regulating cucumber fruit development and ripening and its transcriptional regulation. *Plant Physiol. Bioch.* 64, 70–79. doi: 10.1016/j.plaphy.2012.12.015

Wang, H. S., Yu, C., Fan, P. P., Bao, B. F., Li, T., and Zhu, Z. J. (2015). Identification of two cucumber putative silicon transporter genes in *Cucumis sativus*. *J. Plant Growth Regul.* 34, 332–338. doi: 10.1007/s00344-014-9466-5

Wang, X., Gao, D. L., Sun, J. J., Liu, M., Lun, Y. Y., Zheng, J. S., et al. (2016). An exon skipping in a SEPALLATA-Like gene is associated with perturbed floral and fruits development in cucumber. *J. Integr. Plant Biol.* 58, 766–771. doi: 10.1111/jipb.12472

Weigel,, and Meyerowitz, (1994). The ABCs of floral homeotic genes. *Cell* 79, 180–180. doi: 10.1016/0092-8674(94)90291-7

Weng, Y., Colle, M., Wang, Y., Yang, L., Rubinstein, M., Sherman, A., et al. (2015). QTL mapping in multiple populations and development stages reveals dynamic quantitative trait loci for fruit size in cucumbers of different market classes. *Theor. Appl. Genet.* 128, 1747–1763. doi: 10.1007/s00122-015-2544-7

Yang, Y. Z., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* 33, 47–59. doi: 10.1046/j.0960-7412.2003.01473.x

Zhang, X. L., Zhou, Y., Ding, L., Wu, Z. G., Liu, R. Y., and Meyerowitz, E. M. (2013). Transcription repressor HANABA TARANU controls flower development by integrating the actions of multiple hormones, floral organ specification genes, and GATA3 family genes in *Arabidopsis*. *Plant Cell* 25, 83–101. doi: 10.1105/tpc.112.107854

Construction of a High-Density Genetic Map and Analysis of Seed-Related Traits Using Specific Length Amplified Fragment Sequencing for *Cucurbita maxima*

Yunli Wang [1,2†], **Chaojie Wang** [1,2†], **Hongyu Han** [1,2], **Yusong Luo** [1,2], **Zhichao Wang** [1,2], **Chundong Yan** [1,2], **Wenlong Xu** [1,2] **and Shuping Qu** [1,2*]

[1] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Northeast Region), Ministry of Agriculture and Rural Affairs/Northeast Agricultural University, Harbin, China, [2] College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, China

Seed traits are agronomically important for *Cucurbita* breeding, but the genes controlling seed size, seed weight and seed number have not been mapped in *Cucurbita maxima* (*C. maxima*). In this study, 100 F_2 individual derived from two parental lines, "2013-12" and "9-6", were applied to construct a 3,376.87-cM genetic map containing 20 linkage groups (LGs) with an average genetic distance of 0.47 cM using a total of 8,406 specific length amplified fragment (SLAF) markers in *C. maxima*. Ten quantitative trait loci (QTLs) of seed width (SW), seed length (SL) and hundred-seed weight (HSW) were identified using the composite interval mapping (CIM) method. The QTLs affecting SW, SL and HSW explained a maximum of 38.6%, 28.9% and 17.2% of the phenotypic variation and were detected in LG6, LG6 and LG17, respectively. To validate these results, an additional 150 F_2 individuals were used for QTL mapping of SW and SL with cleaved amplified polymorphic sequence (CAPS) markers. We found that two major QTLs, SL6-1 and SW6-1, could be detected in both SLAF-seq and CAPS markers in an overlapped region. Based on gene annotation and non-synonymous single-nucleotide polymorphisms (SNPs) in the major SWand SL-associated regions, we found that two genes encoding a VQ motif and an E3 ubiquitin-protein ligase may be candidate genes influencing SL, while an F-box and leucinerich repeat (LRR) domain-containing protein is the potential regulator for SW in C. maxima. This study provides the first high-density linkage map of *C. maxima* using SNPs developed by SLAF-seq technology, which is a powerful tool for associated mapping of important agronomic traits, map-based gene cloning and marker-assisted selection (MAS)-based breeding in *C. maxima*.

Keywords: *Cucurbita maxima*, high-density genetic map, specific length amplified fragment sequencing, seed-related traits, cleaved amplified polymorphic sequence

INTRODUCTION

The squash *Cucurbita maxima* Duch (2x = 2n = 40) is an important cucurbitaceous plant that is widely grown worldwide as a commercial crop. It has high nutritional value and health-protective properties and thus has received increasing interest and popularity in breeding. Research on squash genetics and molecular biology has also progressed rapidly in recent years.

A high-density genetic map is not only a key resource for studies on genome structure and genetic relationships but also provides the basis for quantitative trait locus (QTL) mapping and marker-assisted selection (MAS) based on the numbers of polymorphic markers (Wang et al., 2011). Since the first genetic linkage map of the genus *Cucurbita* was constructed with isozymes in 1986 (Weeden and Robinson, 1986), several genetic maps have been developed based on molecular markers. The genetic maps of *Cucurbita pepo* have been constructed with random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), sequence characterized amplified region (SCAR), and single-nucleotide polymorphism (SNP) markers and spanned 1954–2817.6 centimorgans (cM) (Brown and Myers, 2002; Zraidi et al., 2007; Gong et al., 2008a; Esteras et al., 2012; Monteropau et al., 2017). The high-density genetic map of *C. pepo* contained 7,718 markers with an average genetic distance between markers of 0.4 cM (Monteropau et al., 2017). The genetic maps of *Cucurbita moschata* were developed using SSR and SNP markers and spanned 1445.4–3087.03 cM (Gong et al., 2008b; Zhong et al., 2017). The high-density genetic map of *C. moschata* contained 3470 markers with an average genetic distance of 0.89 cM (Zhong et al., 2017). Compared with those of *C. pepo* and *C. moschata*, few maps of *C. maxima* with a limited number of markers have been published. The maps of *C. maxima* spanned 991.5–2566.8 cM, and the latest genetic map contained 458 markers with an average marker density of 5.60 cM (Singh et al., 2011; Ge et al., 2015; Zhang et al., 2015a). A genetic map with a low density of markers has limited application for further fine QTL mapping and MAS in squash breeding; therefore, a high-density genetic map of *C. maxima* is urgently needed.

Recently, a new high-throughput strategy for *de novo* SNP discovery called specific length amplified fragment sequencing (SLAF-seq) has been developed and successfully applied to construct high-density maps of grape (Wang et al., 2017a), gingko (Liu et al., 2017), pear (Wang et al., 2017b), tobacco (Gong et al., 2016), matsudana (Zhang et al., 2016), cauliflower (Zhao et al., 2016), soybean (Qi et al., 2014), sesame (Zhang et al., 2013), watermelon (Shang et al., 2016), and cucumber (Xu et al., 2015). This strategy has also been applied for QTL mapping of important economic and nutritional traits in sesame (Mei et al., 2017), soybean (Li et al., 2017), spinach (Qian et al., 2017), cucumber (Zhu et al., 2016), wax gourd (Jiang et al., 2015), and pumpkin (Zhong et al., 2017). However, SLAF-seq has not been successfully applied in squash. Because most genomes have characteristics of genetic variation and stability (Sun et al., 2013), SNPs are considered the most useful and conventional of all developed molecular markers. High-throughput SNP

markers have potential applications for successful QTL mapping and gene cloning in *C. pepo* (Holdsworth et al., 2016; Capuozzo et al., 2017), *C. moschata* (Zhong et al., 2017) and *C. maxima* (Zhang et al., 2015b).

Seed traits are important agronomic traits for squash breeders. Larger seeds can provide more nutrients for seedlings, which can result in increased resilience against environmental stress. Many artificial-selection-based studies of squash breeding have focused on increasing seed size and weight. Most published reports related to differences in seed size have focused on the diversity and evaluation of the combining ability (Nagar et al., 2017; Darrudi et al., 2018), whereas few reports have focused on gene locations. Four seed width (SW)-associated QTLs were located using 193 F_2 *C. maxima* individuals with logarithm of odds (LOD) values ranging from 3.49 to 4.22 and percent variance explained (PVE) values ranging from 2.87 to 29.68 cM (Tan et al., 2013). However, genes controlling seed length (SL), hundred-seed weight (HSW), and seed number per fruit (SNF) have not been mapped.

The QTLs of important seed traits in *C. maxima* have not been cloned. Therefore, a high-density linkage map with more informative and high-throughput markers is urgently needed. To construct a high-density SNP map and identify QTLs for significant seed traits in *C. maxima*, we used SLAF-seq technology for genotyping a population with 100 F_2 individuals derived from two inbred lines with different agronomic traits. With this map, QTLs associated with SW, SL, SNF and HSW were successfully identified in narrow candidate regions. To locate genes controlling SL and SW, QTL mapping with cleaved amplified polymorphic sequence (CAPS) markers was performed under different growing conditions and in different populations over a two-year period. Then, the coding genes mapped in the candidate region were predicted.

MATERIALS AND METHODS

Plant Material and DNA Extraction

An F_2 mapping population with 100 individuals was derived from a cross between the high-generation inbred lines "2013-12" and "9-6" for SLAF-seq. An additional 150 F_2 individuals were self-pollinated to contrast with the $F_{2:3}$ family for locus verification using CAPS markers. Seedlings of the parental F_2 population and $F_{2:3}$ (10 plants per family) individuals were planted in a greenhouse at Xiangyang Base in Heilongjiang Province. Young leaves from the parents and F_2 populations were collected for DNA extraction with the cetyltrimethylammonium bromide (CTAB) method. All fruits of the F_2 and $F_{2:3}$ individuals were harvested 50 days after pollination, and the seeds were cleaned and dried for seed-related phenotype analysis. The SL, SW, SNF, and HSW were measured for each offspring.

SLAF Library Construction and High-Throughput Sequencing

An improved SLAF-seq strategy was utilized in our experiment. The *C. maxima* genome of inbred the *Cucurbita* line *C. maxima*

cv. Rimu (Sun et al., 2017) was used to design SLAF markers using different enzymes. Then, a pilot SLAF experiment was performed, and the SLAF library was constructed in accordance with the predesigned scheme SLAF Predict (Peking Biomarker Technologies Corporation, Peking, China). For the F_2 population in our study, two enzymes (HaeIII+Hpy166II) (New England Biolabs, NEB, USA) were selected as the most appropriate enzymes to digest the genomic DNA. A single-nucleotide (A) overhang was subsequently added to the digested fragments using the Klenow fragment (3´→5´ exo⁻) (NEB) and dATP at 37°C. Duplex tag-labeled sequencing adapters were ligated to the A-tailed fragments using T4 DNA ligase (ThermoFisher Scientific, MA, USA). PCR was carried out using 25-µl samples containing 100 ng of restriction-ligation DNA samples, 1× reaction buffer, 200 µM dNTPs, 0.5 U of Q5 High-Fidelity DNA Polymerase and 0.5 µM PCR primers (forward primer: 5'-AATGATACGGCGACCACCGA-3', reverse primer: 5'-CAAGCAGAAGACGGCATACG-3') (PAGE-purified, Life Technologies, China). The PCR program was as follows: 94°C for 30 s; 12 cycles of 94°C for 40 s, 65°C for 30 s, and 72°C for 30 s; and 72°C for 5 min. The products were purified using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) and pooled. Next, fragments ranging from 264 bp to 364 bp in length were purified using the QIAquick Gel Extraction Kit (Qiagen, Germany). Finally, the gel-purified products were diluted to 100 ng/µl, and paired-end sequencing was performed on the Illumina HiSeq 2500 system (Illumina, Inc.; San Diego, CA, USA) according to the manufacturer's recommendations.

Sequence Data Grouping and Genotyping

SLAF marker identification and genotyping were performed using procedures described by Sun et al. (2013). Briefly, low-quality reads (quality score < 20e) were filtered out, and then, the raw reads were sorted to each progeny according to the duplex barcode sequences. After the barcodes and terminal 5-bp positions were trimmed from each high-quality read, clean reads from the same sample filtered by SOAP software (Li et al., 2008) were mapped onto the *C. maxima* genome sequence (Sun et al., 2017). Sequences that mapped to the same position were defined as one SLAF locus (Zhang et al., 2015b). Then, SNP loci for each SLAF locus were compared between the parents; first, SLAF loci with more than 3 SNPs were filtered out. Alleles of each SLAF locus were defined according to parental reads with a sequencing depth >10-fold, whereas for each offspring, reads with sequencing depth >1-fold were used to define alleles. For pumpkin, one SLAF locus can contain at most 4 genotypes, so SLAF loci with more than four alleles were discarded subsequently. Only SLAFs with two to four alleles were identified as polymorphic markers and considered potential markers. The polymorphic SLAF loci were genotyped to ensure consistency between the parental and offspring SNP loci. Polymorphic markers were classified into eight segregation patterns. Only one segregation type (aa×bb) was used to construct the genetic map because the polymorphic markers were analyzed according to the F_2 population type. Genotype scoring was performed using a Bayesian approach to further

ensure the genotyping quality (Sun et al., 2013). Low-quality markers for each marker and each individual were counted, and the worst marker or individual was deleted. The chi-square test was performed to examine the segregation distortion, and markers with significant segregation distortion ($P < 0.05$) were excluded from the map construction.

Linkage Map Construction and Evaluation

All high-quality SLAF markers were allocated into 20 linkage groups (LGs) based on their genomic locations. Then, the modified LOD (MLOD) scores between markers were calculated to further confirm the robustness of markers for each LG, and markers with MLOD scores < 5 were filtered prior to ordering. To ensure efficient construction of the high-density and high-quality map, a newly developed HighMap strategy was utilized to order the SLAF markers and correct genotyping errors within the LGs (Liu et al., 2014). First, recombinant frequencies and LOD scores were calculated using a two-point analysis and applied to infer linkage phases. Then, enhanced Gibbs sampling, spatial sampling, and simulated annealing algorithms were combined for iterative marker ordering (Jansen et al., 2001; Van, 2011). The updated recombination frequencies were used to integrate the two parental maps, which optimized the map order in the subsequent simulated annealing cycle. A stable map order was obtained after 3–4 cycles, resulting in a high-quality map including 20 LGs. The SMOOTH error correction strategy was performed according to parental contribution to the genotypes (Os et al., 2005), and a k-nearest neighbor algorithm was applied to impute missing genotypes (Huang et al., 2011). Skewed markers were added to this map by applying a multipoint maximum likelihood method. Map distances were estimated using the Kosambi mapping function (Kosambi, 1944). Marker pairs with zero recombination in each LG had the same genetic distance. The data analysis script draw haplotype-map.pl (Peking Biomarker Technologies Corporation, Peking, China) was used to construct the haplotype map, and draw heatmap.pl (Peking Biomarker Technologies Corporation, Peking, China) was used to construct the heat map.

Relationship Between the Genetic and Physical Maps

The physical positions of the SNPs were determined based on alignment with the reference genome sequence of *C. maxima* (Sun et al., 2017). The collinearity between the genetic and physical positions was determined by plotting each marker's genetic position (in cM) against its physical position (in Mb) using Excel 2007 (Microsoft Corporation, WA, USA). Spearman's correlation coefficients were calculated using the Statistical Analysis System (SAS) program (ND Times, Peking, China).

CAPS of DNA Polymorphisms

To verify the QTL mapping results by SLAF-seq, CAPS markers surrounding the positioning region of the major QTLs associated with SL and SW were selected (at a physical distance from 1 Mb to 5 Mb in chromosome 6). Five hundred-bp sequences

surrounding SNPs from the SLAF-seq data were used to design primers for CAPS markers. The PCR products were digested by a restriction endonuclease (i.e., *EcoRI, HindIII, PstI, ScaI, BamHI, XhoI*) (Thermo Scientific, MA, USA) to verify polymorphisms of the CAPS markers. The primers were designed using Primer Premier 6.0 software with the appropriate CAPS candidate sequences.

The PCR mixture for CAPS amplification contained 20 ng of plant genomic DNA, 10 pmol of the primers, 0.25 mM dNTPs, $10\times$ Taq buffer, and 1 unit of Taq polymerase in a total volume of 20 µl. Touchdown PCR was performed for 7 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 60°C with stepwise decreases of 0.5°C for each cycle, and 60 s at 72°C; 10 cycles of 30 s at 94°C, 30 s at 45°C, and 60 s at 72°C; and post-heating for 7 min at 72°C. The reaction mixture for enzyme digestion contained 5 µl of the PCR product, 3.7 µl of ddH$_2$O, 0.3 µl of the restriction enzyme (10 U/µl) and 1 µl of $10\times$ enzyme buffer, which was incubated at 37°C for 3 h. The enzyme-digested products were examined *via* 2% agarose gel electrophoresis. The primer sequences for the 11 polymorphic CAPS markers are listed in **Table S1**. The enzyme-digested products were examined *via* 2% agarose gel electrophoresis at 150 V for 30 min. The agarose gel electrophoresis results were photographed using the ChampGel 6000 gel documentation and image analysis system.

Phenotyping and QTL Mapping of Seed-Related Traits

The average SL and SW were measured for 10 randomly chosen seeds from each line with three replications, and SNF and HSW data were collected from 100 F$_2$ individuals in 2017 for QTL analysis with SLAF-seq. SL and SW were measured for each individual of an additional 150 F$_2$ individuals in 2017. SL and SW of 10 plants per F$_{2:3}$ family in 2018 were measured. The means of SL and SW within each F$_{2:3}$ family in 2018 were calculated and subjected to QTL analysis with CAPS markers.

The genotype of each mapped markers in 100 F$_2$ individuals was analyzed and markers with duplicate genotypes were removed. The R package ASMap (Taylor and Butler, 2017) was used to map QTLs, and QTL analyses were then performed using the R/qtl package with the composite interval mapping (CIM) model (Churchill and Doerge, 1994). The significance of each QTL interval was tested by a likelihood-ratio statistic (LOD) (Van and Kyazma, 2004). For each trait, the LOD threshold for declaring significant QTLs was established separately with 1000 permutation tests ($P = 0.05$), which ranged from 2.8 to 3 for the various traits. To be conservative, an LOD score of 3 was used for the QTL detection of all traits. The QTLs were named according to the trait name and chromosome.

RESULTS

Characteristics Between the Two Parents

The crossing parents "2013-12" and "9-6" differed in several traits, including floral sex, fruit shape, fruit color, flesh thickness, flesh color, SL, SW, SNF, and HSW (**Figure 1** and **Table S2**). The female parent "2013-12" was subgynoecious with an orange-red fruit color, orange-yellow flesh, and small and light seeds. The male parent "9-6" was androecious with a gray fruit color, light-yellow flesh, and large and heavy seeds. The SL (20.35 mm) and SW (14.36 mm) of the F$_1$ plants were similar to those of the parental "9-6" line, but the SNF (189) and HSW (36.97 g) of F$_1$ were greater than those of the parental lines, with transgressions of 63.36% and 39.25%, respectively. Phenotypic data of SL, SW, SNF, and HSW were collected using 100 F$_2$ individuals in 2017 for QTL mapping with SLAF-seq. Phenotypic data of SL and SW were collected using 150 F$_2$ populations in 2017 and 150 F$_{2:3}$ individuals in 2018 for QTL analysis with CAPS markers. Detailed phenotypic and genotype data of the F$_2$ and F$_{2:3}$ populations are presented in **Tables S2** and **S3**. Seed traits in 100 F$_2$ individuals present continuity variance and follow the normal distribution. According to seed trait data from the parents and 100 F$_2$ individuals, correlation analysis showed that SL was significantly correlated with SW ($r = 0.627$), which indicated that long seeds always appeared together with wide seeds.

SLAF-Seq Data and SNP Markers

DNA sequencing generated a total of 100.4 Gb of raw bases. Each read was ~100 bp in length and 473.35 Mb paired-end reads were obtained after SLAF library construction and high-throughput sequencing. Of these reads, 28.81 Mb reads were from the male parent, 29.41 Mb were from the female parent, and an average of 4.15 Mb were from the 100 F$_2$ individuals (**Table 1**). The raw data of the two parents and 100 F$_2$ individuals have been deposited to the National Center for Biotechnology Information (NCBI) and can be accessed in the Short Read Archive (SRA) sequence database under accession number PRJNA549786. Among these reads, 95.08% of the bases were of high quality, with quality scores of at least 30 (Q30). The average GC content was 40.83%, which suggested a normal distribution, and no AT or GC segregation was found in the reads. The sequencing depth of the 100 F$_2$ individuals ranged from 1.61- to 6.55-fold with an average sequencing depth of 3.20-fold (**Figure S1**), which revealed that the sequencing results were reliable for marker exploration.

A total of 584,994 SLAFs were developed, of which 195,305 showed polymorphisms, with a polymorphism rate of 33.39% (**Table S4**). The number of SLAF markers per LG ranged from 17,334 (LG8) to 43,545 (LG4). A total of 178,376 high-quality SLAFs were detected in the offspring, of which the average sequencing depth was 29.01. In the comparison of the sequencing data from the two parents and offspring, a total of 1,437,901 SNPs were identified. Among these SNPs, the numbers of SNPs in "2013-12" and "9-6" were 690,456 and 587,028, respectively, and the average number of SNPs in each F$_2$ individual was 160,417 (**Table 2**). The number of SNP markers per LG ranged from 29,881 to 77,881. SNPs with the same nucleotides for each allele depth were termed homo-SNPs. A total of 1,197,388 homo-SNPs were classified into eight segregation patterns following the CP group data format of Joinmap4.1 software (Van, 2011) and genotype encoding rules (Zhu, 2015) (**Figure S2**). The 365,639 SNPs with an aa×bb segregation pattern in the F$_2$ population were selected to
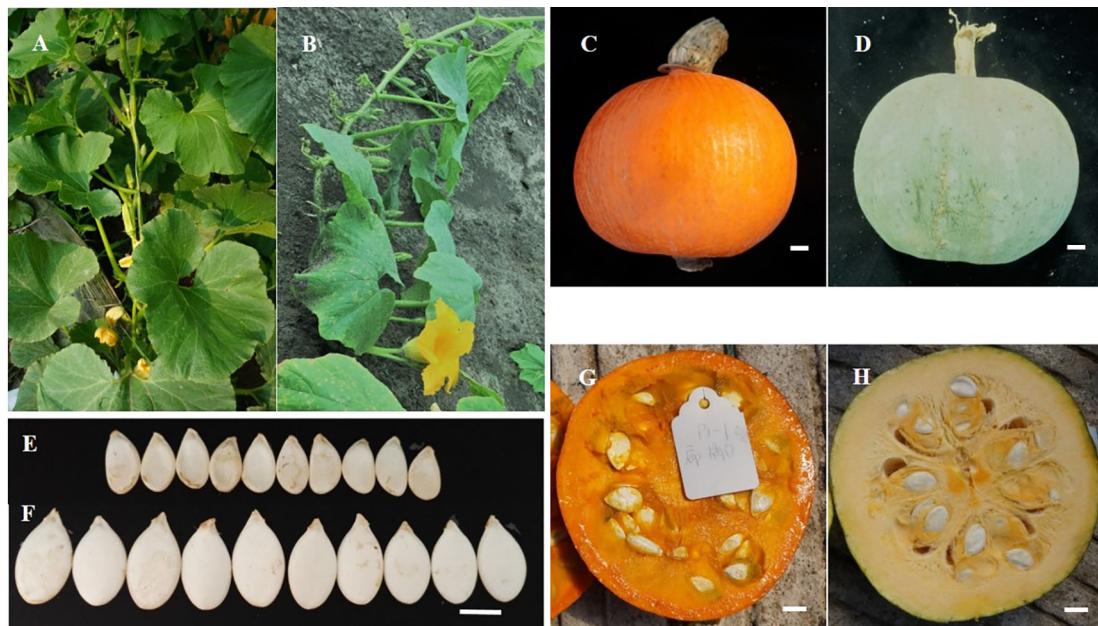
FIGURE 1 | Different phenotype of two parents. A, C, E, and G are the plant type, fruit and seed phenotype of "2013-12". B, D, F, and H are the plant, fruit and seed phenotype of "9-6". **(A)** "2013-12" is subgynoecious type; **(B)** "9-6" is androecious type; **(C)** "2013-12" has orange-red fruits; **(D)** "9-6" has gray fruits; **(E)** "2013-12" has small and light seeds; **(F)** "9-6" has big and heavy seeds; **(G)** "2013-12" shows orang-yellow and thin flesh; **(H)** "9-6" shows light-yellow and thick flesh.

TABLE 1 | Summary of the high-throughput sequencing data.

Sample ID	Total Reads (Mb)	Total Base (Gb)	Q30 (%)	GC (%)
2013-12	28.81	8.62	92.15	38.65
9-6	29.41	8.80	91.76	38.47
Offspring	4.15	0.83	95.14	40.88
Total	473.35	100.43	95.08	40.83

Total reads, total bases, Q30 and GC of sequencing samples are shown.

construct a linkage map because the F_2 population was derived from a cross between two homozygous parents with genotype aa or bb.

Genetic Mapping

Finally, 8,622 markers were used to construct the genetic map based on a sequencing depth in the parents of more than 10-fold, the segregation distortion criteria (P < 0.01), and less than 25% missing marker data in the F_2 population. After removing low-collinearity markers confirmed by calculating MLOD scores between neighboring markers, 8,406 of the 8,622 SNP markers

TABLE 2 | Statistics of SNP number.

Sample ID	SNP number	Heter-SNP number	Homo-SNP number	Heter-ratio
2013-12	690,456	233,696	456,760	33.84%
9-6	587,028	62,459	524,569	10.63%
Offspring	160,417	22,289	138,128	13.89%
Total	1,437,901	318,444	1,119,457	28.45%

SNP numbers, heter-SNP numbers, homo-SNP numbers of samples and heter-ratios of sequencing samples are shown.

were mapped onto 20 LGs (**Table 3**, **Figure S3**). The 8,406 marker sequence and genotypes of markers in F_2 individuals are presented in **Table S5** and **Table S6**. The genetic map spanned a total of 3376.87 cM, with an average distance of 0.47 cM between adjacent markers. The average number of markers in each LG was 420.3, and markers spanned an average length of 168.84 cM. The number of mapped SNP markers ranged from 134 (LG16) to 865 (LG11), and the average distance ranged from 0.21 cM (LG16) to 0.99 cM (LG11). LG12, which was the largest LG (196.35 cM genetic length), contained 490 markers and had an average marker density of 0.40 cM. LG9, which was the smallest LG (121.95 cM genetic length), contained 241 markers and had an average marker density of 0.51 cM. On the genetic map, the largest gap was 11.49 cM and was located in LG14. Gaps <5 cM constituted 99.30% of the total LGs, which showed good uniformity in the distribution.

The total number of reads for the markers used to construct the genetic maps in the two parents were 134,754 and 148,231, and the average depths were 16.03 and 17.63, respectively (**Table S7**). The total marker depth in the F_2 population ranged from 24,382 to 428,849, with a depth of 216,408 in each offspring. The average marker depth in the F_2 population ranged from 6.11 to 51.13, with a depth of 26.24 in each offspring (**Table S8**). The high marker depth could improve the accuracy of marker locations on the genetic map.

Evaluation of the Genetic Map

Haplotype and heat maps were used to evaluate the quality of the genetic map. Potential genotype and marker order errors can be

TABLE 3 | Basic characteristics of 20 LGs.

Linkage Group ID	Total Marker	Total Distance (cM)	Average Distance (cM)	Max Gap (cM)	Gap < 5 cM (%)
LG 01	555	187.33	0.34	10.47	99.46
LG 02	321	151.20	0.47	7.88	98.75
LG 03	430	169.59	0.40	5.42	99.53
LG 04	808	196.01	0.24	6.11	99.88
LG 05	345	171.91	0.50	6.64	99.13
LG 06	542	190.00	0.35	8.17	99.08
LG 07	280	135.77	0.49	6.55	99.64
LG 08	398	169.45	0.43	6.64	99.50
LG 09	241	121.95	0.51	7.48	99.58
LG 10	538	189.32	0.35	3.71	100.00
LG 11	865	180.77	0.21	3.71	100.00
LG 12	490	196.35	0.40	4.81	100.00
LG 13	340	170.73	0.50	5.54	99.71
LG 14	304	142.39	0.47	11.49	98.68
LG 15	305	175.14	0.58	7.86	98.68
LG 16	134	132.10	0.99	8.75	96.99
LG 17	349	193.60	0.56	8.5	99.43
LG 18	683	189.06	0.28	2.71	100.00
LG 19	245	141.37	0.58	6.03	99.18
LG 20	233	172.83	0.74	10.14	98.71
Total	8406	3376.87	0.47	11.49	99.30

Total SNP markers, total distances, average distances, maxs gap and Gap < 5 cM in each linkage group are shown. Numbers of the C. maxima in each LG are also shown. Gap < 5 indicated the percentages of gaps in which the distance between adjacent markers is smaller than 5 cM.

reflected by a haplotype map of the genetic map. Haplotype maps were generated for each F_2 individual and the parental controls using SNP markers. More double crossovers indicated more genotype and marker order errors. As shown in **Figure S4**, most regions in the haplotype map of all F_2 individuals had a common origin, which indicated that a high-quality genetic map was constructed with the 8,406 SNPs. The relationship of recombination between markers from one LG was used to determine the potential ordering errors of the markers. Based on the heat map of the 20 LGs, a strong linkage relationship between two adjoining markers in the 20 LGs was distinctly visible (**Figure S5**). As the distance between the two markers increased, the linkage relationship between them weakened, which illustrated the correct order of markers in the LGs.

The correlation of genetic and physical positions is an important factor for evaluation of genetic maps. Spearman's correlation coefficients between the genetic and physical maps for each LG were calculated to analyze the collinear relationships between the maps (**Table S9**). Spearman's correlation coefficient of each LG ranged from 0.98 to 1, which indicated high collinearity of the genetic and physical maps. As shown in **Figure S6**, marker arrangement on the genetic map was highly consistent with that on the physical map, which indicated good collinearity of the maps and high accuracy of genetic recombination on the genetic map.

QTL Analysis

Among 8,406 markers, 2,041 markers with different genotypes were used for QTL mapping. The LGs and positions in the LGs of 2,041 markers are listed in **Table S10**. According to the genotype of the F_2 population from the SLAF-seq data, the QTLs for SL, SW and HSW were distributed at 10 positions in six LGs, while no significant QTLs were detected in 20 LGs for SNF. The seed traits, LGs, position intervals, starting markers, ending markers, peak markers, LODs, PVEs, additive values, and dominance values are shown in **Table 4** and **Figure 2**. QTLs that could explain more than 15% of the PVEs were considered major-effect QTLs. Four QTLs were detected for SL and were located in four LGs (LG4, LG6, LG17, and LG18). These QTLs explained 7.0-38.6% of the variance. The minor-effect QTLs *SL4-1*, *SL17-1*, and *SL18-1* together explained 30.7% of the variance. The major-effect QTL *SL6-1* was situated between markers Marker157285 and Marker158784 (from 37.9 cM to 42.2 cM) with a peak at Marker158406, and explained 38.6% of the variance with an LOD value of 10.6. Four QTLs for SW were distributed over LG4, LG5, LG6, and LG8 and explained 6.9%–28.9% of the variance. The major-effect QTL *SW6-1* was situated between markers Marker154765 and Marker156546 (from 28.4 cM to 33 cM) with a peak at marker Marker155361 and explained 28.9% of the variance with an LOD value of 7.9. Three minor-effect QTLs, *SW4-1*, *SW5-1*, and *SW8-1*, together explained 30.8% of the variance. Another two QTLs for HSW were located on LG6 and LG17. One major-effect QTL, *HSW17-1*, was situated between markers Marker424927 and Marker424340 (from 99.6 cM to 111.9 cM) with a peak at marker Marker424492, and explained

TABLE 4 | Genetic mapping and QTL analysis of 4 seed-related traits in 100 F_2 individuals.

Trait	QTL	LG	Position interval (cM)	Start marker	End marker	Peak marker	LOD	PVE (%)	Add	Dom
Seed length	*SL4-1*	4	92.5-101.6	Marker93772	Marker94339	Marker93889	4.1	12.6	1.1	0.3
	SL6-1	6	37.9-42.2	Marker157285	Marker158784	Marker158406	10.6	38.6	1.6	1.3
	SL17-1	17	134.9-140.8	Marker421850	Marker421215	Marker421617	3.7	11.1	-0.1	-1.5
	SL18-1	18	143.6-176.2	Marker477731	Marker486042	Marker483115	3.0	7.0	-0.6	-0.4
Seed width	*SW4-1*	4	92.5-122.9	Marker93772	Marker96127	Marker94339	3.5	6.9	0.4	0.6
	SW5-1	5	60.3-88.7	Marker137649	Marker144920	Marker140172	3.1	10.9	0.6	0.5
	SW6-1	6	28.4-33.0	Marker154765	Marker156546	Marker155361	7.9	28.9	1.2	0.3
	SW8-1	8	39.6-46.9	Marker200328	Marker201237	Marker200805	3.9	13.0	0.7	0.5
Hundred seed weight	*HSW6-1*	6	53.5-64.2	Marker161001	Marker163399	Marker162384	3.9	13.2	4.2	0.6
	HSW17-1	17	99.6-111.9	Marker424927	Marker424340	Marker424492	5.4	17.2	0.6	-6.7

The seed traits, linkage groups, position intervals, starting markers, ending markers, peak markers, LODs, PVEs, additive values and dominance values are shown. QTLs of SL, SW, SNF and HSW are shown. LG, Linkage group; PVE, phenotype variance explained; Add, additive value; Dom, dominance value.
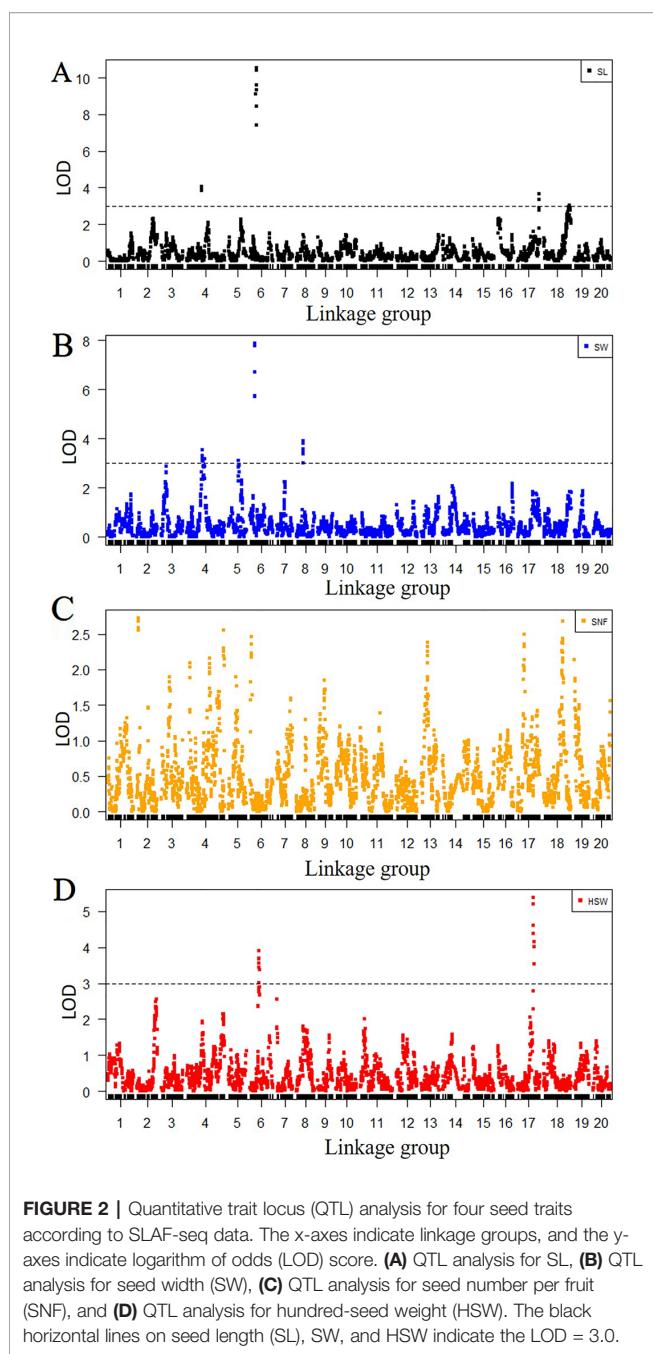
FIGURE 2 | Quantitative trait locus (QTL) analysis for four seed traits according to SLAF-seq data. The x-axes indicate linkage groups, and the y-axes indicate logarithm of odds (LOD) score. **(A)** QTL analysis for SL, **(B)** QTL analysis for seed width (SW), **(C)** QTL analysis for seed number per fruit (SNF), and **(D)** QTL analysis for hundred-seed weight (HSW). The black horizontal lines on seed length (SL), SW, and HSW indicate the LOD = 3.0.

17.2 of the variance with an LOD value of 5.4. In addition, *SL6-1*, *SW6-1*, and *HSW6-1* were located at the same LG.

Linkage Map Construction With CAPS Markers

According to the QTL results of SLAF-seq, the major QTL of SL and SW was located at a physical location of 28.4 cM to 42.2 cM in LG6 (with a physical distance of 1.82 Mb to 2.92 Mb). To verify the location results for SL and SW, 15 pairs of CAPS markers were designed from 1.00 Mb to 5.00 Mb in LG6. Eleven pairs of the CAPS marker showed clear, repeating, polymorphic bands for "2013-12", "9-6" and their F_1 progeny. The 500-bp

sequences surrounding SNPs for eleven CAPS markers are presented in **Table S11**. CAPS production bands consistent with "2013-12" were marked as A, those consistent with "9-6" were marked as B, and those consistent with F_1 were marked as H. We genotyped 150 F_2 plants with these 11 markers (**Table S3**); the LOD profiles of the QTLs for the SL and SW traits are illustrated in **Figure 3** and **Table 5**. The QTL *SL6-1F2* was situated between markers M2374213 and M3507675, with an LOD score of 3.6 and a phenotypic variation of 22.2%. The QTL *SW6-1F2* was identified between markers M1468248 and M1929605; this QTL accounted for 20.8% of the total variation and had an LOD score of 4.8. No significant QTLs were detected in the $F_{2:3}$ population for SL and SW, while the peak regions of QTL *SL6-1F2:3* and *SW6-1F2:3* were located in regions similar to those of the peaks of *SL6-1F2* and *SW6-1F2*. The QTL *SL6-1* had an overlapping region with QTLs *SL6-1F2* at 2.52-2.92 Mb in LG6, while QTL *SW6-1* had an overlapping region with QTLs *SL6-1F2* at 1.82-1.93 Mb in LG6. These findings indicated that CAPS markers M2374213 and M3507675 could be used for MAS of the SL trait, and markers M1468248 and M1929605 could be used for MAS of SW traits.

Predicted Genes for Seed Length and Seed Width

To efficiently select predicted genes controlling the SL and SW traits, the coding genes of the SL-associated region (from 2.52 Mb to 2.92 Mb) and SW-associated region (from 1.82 Mb to 1.93 Mb) in LG6 were analyzed. Protein structures and functions could be influenced by a change in a single nucleotide in the coding region. According to the SNPs and InDels from the SLAF-seq data, genes with missense codons between the two parents were selected, and twenty-two genes were identified in the candidate regions of the major QTLs for SL and SW. The gene ID and probable function information are listed in **Table 6**. Fifteen genes were annotated in SL-associated regions, and seven genes were annotated in SW-associated regions. According to seed-traits related gene descriptions (Sun et al., 2017), one gene encoding a VQ motif (*CmaCh06G005530.1*, from 2628780 bp to 2629112 bp in LG6) and one gene encoding E3 ubiquitin-protein ligase (*CmaCh06G005450*, from 2581752 bp to 2588335 bp) with a conserved domain regulates SL in other plant species (Song et al., 2007; Wang et al., 2010). Another gene encoding an F-box and leucine rich repeat (LRR) domains-containing protein (*CmaCh06G004140.1*, from 1917891 bp to 1923829 bp) with a conserved domain regulates SW in several plant species (Luo et al., 2005). The amino acid sequence alignment between "2013-12" and "9-6" for these genes is shown in **Figure S7**. The amino acid sequence differences in *CmaCh06G005450* and *CmaCh06G005530.1* between the two parental lines occurred in a conserved sequence (amino acid residues 5 to 333 aa and 14 to 75 aa, respectively). The amino acid sequence differences in *CmaCh06G004140.1* between the two parental lines occurred in an N-terminal C2 conservative domain (11-140 aa) and a structural maintenance of chromosomes (SMC) -super family conservative domain (276-1026 aa), which could affect protein function. These results indicate that *CmaCh06G005450*,

CmaCh06G005530.1, and *CmaCh06G004140.1* might be good candidates to explain differences in SL and SW between the parents and offspring. Further experiments are needed to confirm the functions of these genes in *Cucurbita*.

DISCUSSION

SL, SW, SNF, and HSW have significant effects on squash yield and seed quality. Breeders prefer high seed yields and heavy seeds. The F_1 plants derived from a cross between the "2013-12" line and "9-6" lines exhibited over-parent heterosis in SNF and HSW. The "9-6" line could provide a new allele source for pumpkin breeding with large and heavy seeds.

High-density genetic maps are a highly valuable tool for map-based cloning and MAS. This study reports the construction of the first high-density linkage map of *C. maxima* using SNPs identified by SLAF-seq technology. When we compared the sequencing data from the two parents and offspring, a total of 584,994 SLAF markers and 1,437,901 SNPs were identified, and the average sequencing depth was 29.01. A number of markers will be beneficial for gene determinations and molecular cloning in pumpkin breeding. The genetic map studied herein spanned a total of 3,376.87 cM, and the linkage map in our study exhibited 20 LGs with approximately the same physical distances as the reference genome (Sun et al., 2017). It was indicated that the genetic map and SLAF markers in our study were of good quality. A total of 8,406 SNPs were used to construct a genetic

map in 20 LGs with an average distance of 0.47 cM between adjacent markers, which meant that more markers were identified here than for *C. maxima* with 458 bin markers (Zhang et al., 2015b), *C. moschata* with 3,470 SNPs (Zhong et al., 2017) and *C. pepo* with 7,718 SNPs (Monteropau et al., 2017). The mean distance between markers in the genetic map in our study was shorter than those for *C. maxima,* with a mean marker density of 5.60 cM (Zhang et al., 2015b), and *C. moschata,* with a mean marker density of 0.89 cM (Zhong et al., 2017), and was almost as short as that for *C. pepo*, with a mean marker density of 0.4 cM (Monteropau et al., 2017). The published high-density genetic maps of *Cucurbita* were constructed using genotyping-by-sequencing (GBS) (Zhang et al., 2015b; Monteropau et al., 2017) and double-digest restriction-associated DNA (ddRAD) libraries (Zhong et al., 2017). GBS and ddRAD genotyping is performed by randomly interrupting genomic DNA with restriction enzymes, whereas SLAF-seq is performed by sequencing the paired ends of sequence-specific restriction fragments (Xu et al., 2015; Zhang et al., 2015b; Zhong et al., 2017). Therefore, SLAF-seq provides better repeatability and deeper sequencing than the above methods.

High-density markers and sites are necessary for map-based cloning of genes. In the genetic map of our study, LG9 was the largest LG (196.35 cM) and contained 490 markers, whereas LG12 was the smallest (121.95 cM) and contained 241 markers. On the map of *C. maxima* published by Zhang et al. (2015b), the smallest LG was LG10 (47.61 cM), which harbored 10 markers
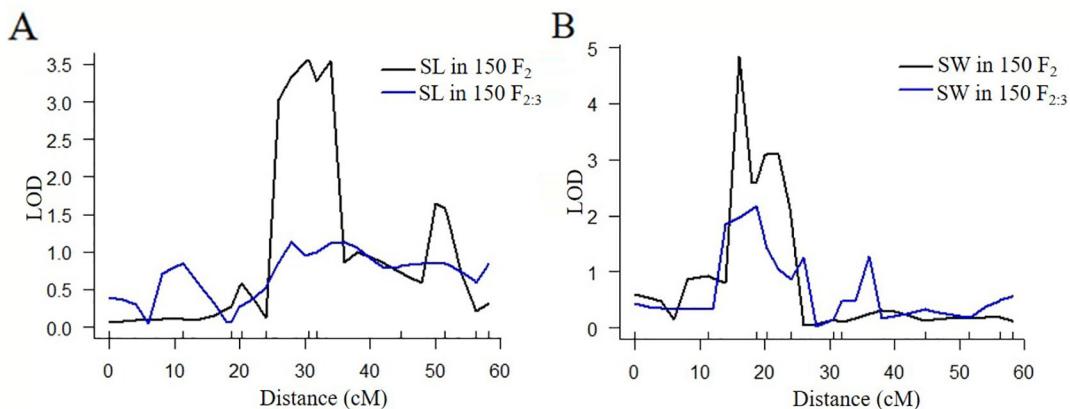
FIGURE 3 | Quantitative trait locus (QTL) analysis for seed length and seed width by CAPS markers. **(A)** QTL analysis for seed length in 150 F_2 and $F_{2:3}$ individuals. **(B)** QTL analysis for seed width in 150 F_2 and $F_{2:3}$ individuals.

TABLE 5 | Position of major QTLs for SL and SW using CAPS markers.

QTL name	Trait	Position interval (cM)	Left marker	Right marker	Peak marker	LOD Threshold	PVE (%)	Add	Dom
SL6-1F2	SL	24.1-44.7	M2374213	M3507675	M2802290	3.6	22.2	-1.2	-0.1
SL6-1F2:3	SL	0-58.2	M1047465	M4930644	M2802290	1.1	14.6	-0.8	0.0
SW6-1F2	SW	11.3-18.8	M1468248	M1929605	M1929605	4.8	20.8	-0.8	0.0
SW6-1F2:3	SW	11.3-44.7	M1468248	M3507675	M1929605	2.2	22.2	-0.5	0.3

QTL names, position intervals, left markers and right markers, peak markers, LOD Thresholds, additive value, dominance value and PVEs are shown. ADD, additive value; DOM, dominance value; PVE, phenotype variance explained. Parent line "2013-12" account for the Add and Dom value.

TABLE 6 | The predicted genes in major QTLs for SL and SW.

QTL name	Trait	Gene ID	Gene description
SL6-1	Seed Length	*CmaCh06G005340.1*	CSL zinc finger domain-containing protein
		CmaCh06G005350.1	MD-2-related lipid recognition domain-containing protein
		CmaCh06G005360.1	protein disulfide oxidoreductase
		CmaCh06G005400.1	Polyketide cyclase/dehydrase and lipid transport superfamily protein
		CmaCh06G005450.1	E3 ubiquitin-protein ligase RHF2A
		CmaCh06G005480.1	Sucrose-6F-phosphate phosphohydrolase
		CmaCh06G005500.1	ABC transporter G family member 22-like
		CmaCh06G005530.1	VQ motif
		CmaCh06G005540.1	Zein-binding
		CmaCh06G005620.1	protein downstream neighbor of Son isoform X1
		CmaCh06G005640.1	homogentisate phytyltransferase 2
		CmaCh06G005780.1	Ferredoxin–NADP reductase
		CmaCh06G005820.1	Beta-carotene isomerase D27
		CmaCh06G005890.1	Serine/threonine protein kinase
		CmaCh06G005920.1	N6-adenosine-methyltransferase MT-A70-like
SW6-1	Seed width	*CmaCh06G003930.1*	Lipid A export ATP-binding/permease protein MsbA
		CmaCh06G003980.1	Xyloglucan endotransglucosylase/hydrolase
		CmaCh06G004000.1	ACT domain-containing family protein
		CmaCh06G004090.1	UDP-Glycosyltransferase superfamily protein
		CmaCh06G004100.1	UDP-Glycosyltransferase superfamily protein
		CmaCh06G004120.1	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain-containing protein
		CmaCh06G004140.1	F-box and Leucine Rich Repeat domains containing protein

Seed traits, gene IDs and gene descriptions of QTL are shown.

on 1 oriented scaffold, and the largest LG was LG4 (268.00 cM), which harbored 41 markers on 2 oriented scaffolds. These two genetic maps of *C. maxima* showed little consistency in LG length. The evenly distributed SNPs and high-density sites studied herein were developed with SLAF-seq and HighMap methods. The haplotype map, heat map and Spearman correlation coefficients of the genetic map further ensured the SNP location accuracy in the LGs.

On the published high-density linkage map of *C. maxima*, the distance between neighboring markers ranged from 0.05 cM to 44.89 cM, and 8 large gaps (≥18 cM) were detected in LG2, LG3, LG6, LG7, LG11, LG15, and LG19 (Zhang et al., 2015b). On the published high-density linkage map of *C. moschata*, the largest interval gap was 22.30 cM, and large gaps were detected in LG8, LG13, LG16, and LG19 (Zhong et al., 2017). The large gaps and low sequencing coverage have limited further fine mapping and MAS in *Cucurbita*. Therefore, additional markers and high-coverage genetic maps are needed. In our study, the largest interval gap was only 11.49 cM (in LG14), and gaps <5 cM constituted up to 99.30% of the total LG, which illustrated

successful construction of a map with improved coverage. There may be two possible explanations for the high-coverage genetic map. First, the two elite parents used for this map presented significant differences in several traits (**Figure 1**), which indicated the existence of great genetic diversity between the parents. Second, to obtain SLAF markers with few repeated sequences that are evenly distributed in 20 LGs, the squash genome was used to design marker discovery experiments using different enzymes. Two enzymes were used to digest the genomic DNA in this experiment, which ensured that SLAF markers were effectively discovered on a genomic map.

Based on differences in seed trait composition between the two parents in our study, we identified a total of ten QTLs for SL, SW and HSW with the CIM method. Four QTLs were found for SL, four QTLs were found for SW, and two QTLs were found for HSW in our study. No significant QTLs for SNF were detected in our study, most likely because of a slight difference between the parents "2013-12" (116 SNF) and "9-6" (105 SNF). In cucumber and watermelon, 12 QTLs were found for SL with PVE values ranging from 2.20% to 28.90%, 12 QTLs for SW with PVE values ranging from 2.24% to 24.10%, and 10 QTLs for HSW with PVE values ranging from 4.50% to 28.30% (Wang et al., 2014; Zhou et al., 2016). In our study, major QTL *SL6-1* was homologous to Chr3, with a physical distance of 28.29 Mb-32.43 Mb in cucumber (Qi et al., 2013), which is near the SSR22874 marker of *SL3.1* (at a physical distance of 34.08 Mb) in cucumber (Wang et al., 2014). These results indicated that SL is controlled by homologous genes among species. Homologous regions of major QTLs for SW and HSW in cucumber or watermelon (Zhou et al., 2016) were located on different chromosomes or in different locations, in contrast to published QTLs in cucumber and watermelon (Wang et al., 2014; Zhou et al., 2016). Thus, multiple genes may control SW and HSW in different cucurbit species.

According to the seed size data, we found that long seeds always appeared together with wide seeds in the offspring in our study (**Table S2**). Further analysis showed that SL was significantly correlated with SW. The QTL of *SL-1* was in a region similar to that of *SW6-1* in the same LG, resulting in similar colocalization between SL and SW. It would benefit breeding for large seeds. In published studies of seed traits, the major QTLs for SL and SW in watermelon were colocalized in the same LG (Meru and McGregor, 2013; Zhou et al., 2016), which was similar to our results. Four significant QTLs for SL and four significant QTLs for SW were identified, and the flanking markers for these QTLs could be used in molecular MAS to increase the seed size in squash seedlings during the selection of hybrid offspring.

Quantitatively inherited traits can be influenced by the environment and mapping population. In our study, we detected seed size traits (seed length and seed width) with 100 F_2 populations by the CIM method from SLAF-seq data in 2017 (the related QTLs were named *SL4-1* to SL17-1 and *SW5-1* to SW8-1), 150 F_2 populations by CAPS markers in 2017 (the QTLs were named *SL6-1F2* and *SW6-1F2*), and 150 $F_{2:3}$ individuals by CAPS markers in 2018 (the QTLs were named *SL6-1F2:3* and *SW6-1F2:3*).The peaks of QTLs *SL6-1F2* and *SL6-1F2:3* were

collected from the same marker, and the peaks of QTLs *SW6-1F2* and *SW6-1F2:3* were also collected from the same marker. The major QTL *SL6-1* mapped with 100 F_2 individuals had an overlapping region with *SL6-1F2* mapped with additional 150 F_2 individuals, indicating that this region had high LOD scores and PVE values in different individuals from the same year. On the other hand, the QTL *SL6-1F2:3* mapped with 150 $F_{2:3}$ individuals had a low LOD value, which might have been influenced by the different growing environments in the different years. Similar conditions were found in SW QTLs. The overlapped region of *SL6-1* and *SL6-1F2* (from 2.52 Mb to 2.92 Mb) was the SL-associated region, and the overlapped region of *SW6-1* and *SW6-1F2* (from 1.82 Mb to1.93 Mb) was the SW-associated region. Closely linked CAPS markers (M2374213, M3507675, M1468248, and M1929605) could be used for MAS-based breeding in *C. maxima* for seed elongation and the radius. With regard to the available SLAF-seq data, crossed parents with differences in multiple phenotypes were used in our study, and the QTLs controlling important agronomic traits were mapped. Only 100 F_2 individuals were used to construct the SLAF library and map complex seed-related traits in *C. maxima*, and accurate QTL mapping results were obtained. Therefore, SLAF-seq is a useful gene mapping technology for small populations and exhibits a high success rate and stability.

The early growth of endosperm is coordinated with the seed coat growth and plays a direct role in determining the ripened seed size (Garcia et al., 2005; Ingram, 2010). The seed size of *Arabidopsis* is regulated by the *IKU-MINI* pathway, which is thought to control endosperm growth. The *IKU-MINI* pathway includes *IKU1*, which encodes a protein containing a VQ motif; and *MINI3,* which encodes a WRKY transcription factor (Wang et al., 2010). *IKU2* also encodes an LRR kinase and shows a recessive mode of action causing reduction in endosperm growth accompanied by precocious cellularization and reduced seed size in *Arabidopsis* (Luo et al., 2005). In the SL-associated region, we found *CmaCh06G005530.1*, which encodes a protein containing a VQ motif. We found one amino acid difference in *CmaCh06G005530.1* between the two parental lines in a conserved sequence (**Figure S7B**). In the SW-associated region, *CmaCh06G004140*, encoding an F-box and LRR domain-containing protein, had several amino acid differences between the two parental lines in a conserved domain of SMC, which controls the cell cycle control and cell division (**Figure S7C**). These findings suggested that *CmaCh06G005530.1* and *CmaCh06G004140* were the candidate genes responsible for seed length and width, and the seed size of *C. maxima* was also regulated by the *IKU1* and *IKU2* pathway. One gene, *GW2*, encoding a protein with E3 ubiquitin ligase activity, was found to affect the grain size, weight and yield by controlling the endosperm size in rice (Song et al., 2007). Loss of *GW2* function could increase cell numbers, resulting in increased spikelet hull size and enhanced grain size. In the SL-associated region, *CmaCh06G005450* was selected, which encodes E3 ubiquitin-protein ligase. One amino acid difference in *CmaCh06G005450* between the two parental lines was observed in a conserved sequence (**Figure S7A**). This finding

indicated that the *CmaCh06G005450* gene might affect seed size, which is thought to control cell numbers in our *Cucurbita* materials. The SNP information from SLAF-seq was limited compared to the whole-genome resequencing results, and some regions within the QTL interval were not sequenced. Coding genes without nonsynonymous SNPs in major QTL regions were also considered candidate genes. CAPS markers M2374213, M3507675, M1468248, and M1929605 were used for single segment substitution line (SSSL) construction. Based on it, further fine mapping of SL and SW genes and gene expression during the seed development stage will be performed in our laboratory.

CONCLUSION

By using 100 F_2 populations from two morphologically diverse parents, a high-density genetic map of *C. maxima* was constructed by SLAF-seq. With this map, 10 QTLs for seed-related traits were identified, and major QTLs for SW, SL, and HSW were detected in *C. maxima* for the first time. An additional 150 F_2 and $F_{2:3}$ populations were used to map SW and SL with CAPS markers, and the results indicated that the QTL mapping using SLAF-seq was accurate. Nonsynonymous SNPs and gene descriptions of the SL- and SW-associated regions suggested that the genes encoding a VQ motif gene, E3 ubiquitin-protein ligase and an F-box and LRR domain-containing protein might be associated with SW and SL in *C. maxima*. In summary, the high-density genetic map studied herein is a valuable tool for association mapping of important agronomic traits, map-based gene cloning and MAS breeding in *Cucurbita*.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the the National Center for Biotechnology Information (NCBI), accession number PRJNA549786.

AUTHOR CONTRIBUTIONS

YW performed data analysis, preparing the manuscript. CW contributed to collecting phenotypic characteristics and DNA extraction. HH contributed to CAPS demonstration test. WX and ZW contributed to growing plants. YL and CY contributed to providing experimental material. SQ, the corresponding author, oversaw all activities related to the project implementation and manuscript development.All authors read and approved the final version of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01782/full#supplementary-material

REFERENCES

Brown, R. N., and Myers, J. (2002). A genetic map of squash (*Cucurbita* sp.) with randomly amplified polymorphic DNA markers and morphological markers. *J. Am. Soc. Hort. Sci.* 127 (4), 568–575. doi: 10.21273/JASHS.127.4.568

Capuozzo, C., Formisano, G., Iovieno, P., Andolfo, G., Tomassoli, L., Barbella, M. M., et al. (2017). Inheritance analysis and identification of SNP markers associated with ZYMV resistance in *Cucurbita pepo*. *Mol. Breed.* 37 (8), 99. doi: 10.1007/s11032-017-0698-5

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1101/gad.8.21.2653

Darrudi, R., Nazeri, V., Soltani, F., Shokrpour, M., and Ercolano, M. R. (2018). Evaluation of combining ability in *Cucurbita pepo* L. and *Cucurbita moschata* Duchesne accessions for fruit and seed quantitative traits. *J. Appl. Res. Med. Aromatic Plants* 9, 70–77. doi: 10.1016/j.jarmap.2018.02.006

Esteras, C., Gomez, P., Monforte, A., Blanca, J., Dolera, N. V., Roig, C., et al. (2012). High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. *BMC Genom.* 13 (1), 80–102. doi: 10.1186/1471-2164-13-80

Garcia, D., Fitz Gerald, J. N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* 17 (1), 52–60. doi: 10.1105/tpc.104.027136

Ge, Y., Li, X., Yang, X. X., Cui, C. S., and Qu, S. P. (2015). Genetic linkage map of *Cucurbita maxima* with molecular and morphological markers. *Genet. Mol. Res.* 14 (2), 5480–5484. doi: 10.4238/2015.May.22.18

Gong, L., Stift, G., Kofler, R., Pachner, M., and Lelley, T. (2008a). Microsatellites for the genus Cucurbita and SSR-based genetic linkage map of *Cucurbita pepo* L. *Theoret. Appl. Genet.* 117 (1), 37–48. doi: 10.1007/s00122-008-0750-2

Gong, L., Pachner, M., Kalai, K., and Lelly, T. (2008b). SSR-based genetic linkage map of *Cucurbita moschata* and its synteny with *Cucurbita pepo*. *Genome* 51 (11), 878–889. doi: 10.1139/G08-072

Gong, D., Huang, L., Xu, X., Wang, C., and Ren, M. (2016). Construction of a high-density SNP genetic map in flue-cured tobacco based on SLAF-seq. *Mol. Breed.* 36 (7), 100. doi: 10.1007/s11032-016-0514-7

Holdsworth, W. L., LaPlant, K. E., Bell, D. C., Jahn, M. M., and Mazourek, M. (2016). Cultivar-based introgression mapping reveals wild species-derived *Pm-0*, the major powdery mildew resistance locus in squash. *Plos One* 11 (12), e0167715. doi: 10.1371/journal.pone.0167715

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., et al. (2011). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genet.* 44 (1), 32–39. doi: 10.1038/ng.1018

Ingram, G. C. (2010). Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* 247 (3-4), 195–214. doi: 10.1007/s00709-010-0184-y

Jansen, J., ADe, J., and Ooijen, J. W. V. (2001). Constructing dense genetic linkage maps. *Theoret. Appl. Genet.* 102 (6-7), 1113–1122. doi: 10.1007/s001220000489

Jiang, B., Liu, W., Xie, D., Peng, Q., He, X., Lin, Y., 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 Genom.* 16 (1), 1035. doi: 10.1186/s12864-015-2220-y

Kosambi, D. (1944). The estimation of map distances from recombination values. *Annals Human Genet.* 12 (1), 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24 (5), 713–714. doi: 10.1093/bioinformatics/btn025

Li, B., Fan, S., Yu, F., Chen, Y., Zhang, S., Han, F., et al. (2017). High–resolution mapping of QTL for fatty acid composition in soybean using specific locus amplified fragment sequencing. *Theoret. Appl. Genet.* 130 (7), 1467–1479. doi: 10.1007/s00122-017-2902-8

Liu, D. Y., Ma, C. X., Hong, W. G., 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

Liu, H., Cao, F., Yin, T., and Chen, Y. (2017). A highly dense genetic map for Ginkgo biloba constructed using sequence-based markers. *Front. Plant Sci.* 8, 1041–1050. doi: 10.3389/fpls.2017.01041

Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., and Chaudhury, A. (2005). *MINISEED3* (*MINI3*), a WRKY family gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (LRR) kinase gene, are regulators of seed size in *Arabidopsis*. *Proceed. Nat. Acad. Sci. U. S. A.* 102 (48), 17531–17536. doi: 10.1073/pnas.0508418102

Mei, H., Liu, Y., Du, Z., Wu, K., Cui, C., Jiang, X., et al. (2017). High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. *Front. Plant Sci.* 8, 636–646. doi: 10.3389/fpls.2017.00636

Meru, G., and McGregor, C. (2013). Genetic mapping of seed traits correlated with seed oil percentage in watermelon. *Hort. Sci.* 48 (8), 955–959. doi: 10.1007/s13580-013-0059-1

Monteropau, J., Blanca, J., Esteras, C., Martínezpérez, E. M., Gómez, P., Monforte, A. J., et al. (2017). An SNP-based saturated genetic map and QTL analysis of fruit-related traits in Zucchini using Genotyping-by-sequencing. *BMC Genom.* 18 (1), 94. doi: 10.1186/s12864-016-3439-y

Nagar, A., Sureja, A. K., Munshi, A. D., Bhardwaj, R., Kumar, S., and Tomar, B. S. (2017). Heritability, correlation and genetic divergence for different seed traits in pumpkin (*Cucurbita moschata*). *Indian J. Agri. Sci.* 87 (11), 1519–1523.

Os, H. V., Stam, P., Visser, R. G. F., and Eck, H. J. V. (2005). SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. *Theoret. Appl. Genet.* 112 (1), 187–194. doi: 10.1007/s00122-005-0124-y

Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nature Genet.* 45, 1510–1515. doi: 10.1038/ng.2801

Qi, Z., Huang, L., Zhu, R., Xin, D., Liu, C., Han, X., et al. (2014). A high-density genetic map for soybean based on specific length amplified fragment sequencing. *Plos One* 9 (8), e104871. doi: 10.1371/journal.pone.0104871

Qian, W., Fan, G. Y., Liu, D. D., Zhang, H. L., Wang, X. W., Wu, J., et al. (2017). Construction of a high-density genetic map and the X/Y sex-determining gene mapping in spinach based on large-scale markers developed by specific-locus amplified fragment sequencing (SLAF-seq). *BMC Genom.* 18 (1), 276–286. doi: 10.1186/s12864-017-3659-9

Shang, J., Li, N., Li, N., Xu, Y., Ma, S., and Wang, J. M. (2016). Construction of a high-density genetic map for watermelon (*Citrulluslanatus* L.) based on large-scale SNP discovery by specific length amplified fragment sequencing (SLAF-seq). *Sci. Hort.* 203, 38–46. doi: 10.1016/j.scienta.2016.03.007

Singh, A. K., Singh, R., Weeden, N. F., Robinson, R. W., and Singh, N. K. (2011). A linkage map for *Cucurbita maxima* based on randomly amplified polymorphic DNA (RAPD) markers. *Indian J. Hort.* 68 (1), 44–50. doi: 10.1016/j.scienta.2011.01.011

Song, X. J., Huang, W., Shi, M., Zhu, M. Z., and Lin, H. X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genet.* 39 (5), 623–630. doi: 10.1038/ng2014

Sun, X. W., Liu, D. Y., Zhang, X. F., Li, W. B., Liu, H., Hong, W. G., 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, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid Cucurbita Genomes. *Mol. Plant* 10 (10), 1293–1306. doi: 10.1016/j.molp.2017.09.003

Tan, X. Z., Ge, Y., Xu, W. L., Cui, C. S., and Qu, S. P. (2013). Construction of genetic linkage map and QTL analysis for seed width in pumpkin (*Cucurbita maxima*). *Acta Bot. Boreali-Occidentalia Sinica* 33 (4), 697–702.

Taylor, J., and Butler, D. (2017). R package ASMap: efficient genetic linkage map construction and diagnosis. *J. Stat. Softw.* 79 (6), 1–29. doi: 10.18637/jss.v079.i06

Van, O. J., and Kyazma, B. V. (2004). MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. *Plant Res. Int.* Wageningen, Netherlands.

Van, O. J. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* 93 (5), 343–349. doi: 10.1017/S0016672311000279

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., et al. (2010). The VQ motif protein IKU1 regulates endosperm growth and seed size in *Arabidopsis*. *Plant J.* 63 (4), 670–679. doi: 10.1111/j.1365-313X.2010.04271.x

Wang, J., Lydiate, D. J., Parkin, I. A., Falentin, C., Delourme, R., Carion, P. W., et al. (2011). Integration of linkage maps for the *Amphidiploid Brassica napus* and comparative mapping with *Arabidopsis* and *Brassica rapa*. *BMC Genom.* 12 (1), 101–121. doi: 10.1186/1471-2164-12-101

Wang, M., Miao, H., Zhang, S. P., Liu, S. L., Dong, S. Y., Wang, Y., et al. (2014). Inheritance analysis and QTL mapping of cucumber seed size. *Acta Hort. Sinica* 41 (1), 63–72. doi: 10.16420/j.issn.0513-353x.2014.01.010

Wang, J., Su, K., Guo, Y., Xing, H., Zhao, Y., Liu, Z., et al. (2017a). Construction of a high-density genetic map for grape using specific length amplified fragment (SLAF) sequencing. *Plos One* 12 (7), e0181728. doi: 10.1371/journal.pone.0181728

Wang, L., Li, X., Wang, L., Xue, H., Wu, J., Yin, H., et al. (2017b). Construction of a high-density genetic linkage map in pear (*Pyrus communis × Pyrus pyrifolia nakai*) using SSRs and SNPs developed by SLAF-seq. *Sci. Hort.* 218, 198–204. doi: 10.1016/j.scienta.2017.02.015

Weeden, N. F., and Robinson, R. W. (1986). Allozyme segregation ratios in the interspecific cross *Cucurbita maxima × C. ecuadorensis* suggest that hybrid breakdown is not caused by minor alteration in chromosome structure. *Genetics* 114 (2), 593–609.

Xu, X., Xu, R., Zhu, B., Yu, T., Qu, W., Lu, L., et al. (2015). A high-density genetic map of cucumber derived from specific length amplified fragment sequencing (SLAF-seq). *Front. Plant Sci.* 5 (768), 768–776. doi: 10.3389/fpls.2014.00768

Zhang, Y., Wang, L., Xin, H., Li, D., Ma, C., Ding, X., et al. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13 (1), 141. doi: 10.1186/1471-2229-13-141

Zhang, J., Zhang, Q. X., Cheng, T. R., Yang, W. R., Pan, H. T., Zhong, J. J., et al. (2015a). 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), 183–192. doi: 10.1093/dnares/dsv003

Zhang, G., Ren, Y., Sun, H., Guo, S., Zhang, F., Zhang, J., et al. (2015b). A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). *BMC Genom.* 16 (1), 1101–1114. doi: 10.1186/s12864-015-2312-8

Zhang, J., Yuan, H., Li, M., Li, Y., Wang, Y., Ma, X. J., et al. (2016). A high-density genetic map of tetraploid *Salix matsudana* using specific length amplified fragment sequencing (SLAF-seq). *Plos One* 11 (6), e0157777. doi: 10.1371/journal.pone.0157777

Zhao, Z., Gu, H., Sheng, X., Yu, H., Wang, J., Huang, L., et al. (2016). Genome-wide single-sucleotide polymorphisms discovery and high-density genetic map construction in cauliflower using specific-locus amplified fragment sequencing. *Front. Plant Sci.* 7, 334. doi: 10.3389/fpls.2016.00334

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

Zhou, H. W., Lu, B. Y., Ma, H. Y., Gao, P., Luan, F. S., Gao, Q. F., et al. (2016). QTL mapping of watermelon seed traits. *Acta Hort. Sinica* 43 (4), 715–723. doi: 10.16420/j.issn.0513-353x.2015-0973

Zhu, W. Y., Huang, L., Chen, L., Yang, J. T., Wu, J. N., Qu, M. L., et al. (2016). A high-density genetic linkage map for cucumber (*Cucumis sativus* L.): based on specific length amplified fragment (SLAF) sequencing and QTL analysis of fruit traits in cucumber. *Front. Plant Sci.* 7 (569), 437. doi: 10.3389/fpls.2016.00437

Zhu, Y. (2015)., in: Construction of genetic linkage maps using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia*), Shandong Agricultural UniversityShandong, China.

Zraidi, A., Pachner, M., Shojaeiyan, A., Gong, L., and Lelley, T. (2007). A consensus map for *Cucurbita pepo*. *Mol. Breed.* 20 (4), 375–388. doi: 10.1007/s11032-007-9098-6

Comparative Transcriptome Analysis Provides Insights Into Yellow Rind Formation and Preliminary Mapping of the *Clyr* (*Yellow Rind*) Gene in Watermelon

Dongming Liu[1,2†], Huihui Yang[1†], Yuxiang Yuan[2], Huayu Zhu[1], Minjuan Zhang[1], Xiaochun Wei[2], Dongling Sun[1], Xiaojuan Wang[1], Shichao Yang[1] and Luming Yang[1*]

[1] College of Horticulture, Henan Agricultural University, Zhengzhou, China, [2] Institute of Horticulture, Henan Academy of Agricultural Sciences, Zhengzhou, China

As an important appearance trait, the rind color of watermelon fruit affects the commodity value and further determines consumption choices. In this study, a comparative transcriptome analysis was conducted to elucidate the genes and pathways involved in the formation of yellow rind fruit in watermelon using a yellow rind inbred line WT4 and a green rind inbred line WM102. A total of 2,362 differentially expressed genes (DEGs) between WT4 and WM102 at three different stages (0, 7, and 14 DAP) were identified and 9,770 DEGs were obtained by comparing the expression level at 7 DAP and 14 DAP with the former stages of WT4. The function enrichment of DEGs revealed a number of pathways and terms in biological processes, cellular components, and molecular functions that were related to plant pigment metabolism, suggesting that there may be a group of common core genes regulating rind color formation. In addition, next-generation sequencing aided bulked-segregant analysis (BSA-seq) of the yellow rind pool and green rind pool selected from an F_2 population revealed that the yellow rind gene (*Clyr*) was mapped on the top end of chromosome 4. Based on the BSA-seq analysis result, *Clyr* was further confined to a region of 91.42 kb by linkage analysis using 1,106 F_2 plants. These results will aid in identifying the key genes and pathways associated with yellow rind formation and elucidating the molecular mechanism of rind color formation in watermelon.

Keywords: watermelon, yellow rind, transcriptome, BSA-seq, gene mapping

INTRODUCTION

Watermelon (*Citrullus lanatus*) is an important horticultural crop in the Cucurbitaceae family and is one of the top ten most consumed fresh fruits globally. As one of the most popular fruits in many countries, more than 117 million tons of watermelon was produced in 2016 according to the latest statistical data from the FAO (http://www.fao.org). The diploid watermelon has 22 chromosomes (2n = 2x = 22) with

an estimated genome size of ~425 Mb (Guo et al., 2013). The draft genome of the East Asia watermelon cultivar 97103 has been sequenced and assembled using NGS sequencing technology (Guo et al., 2013), which greatly facilitated genetic and genomics studies, such as marker development and gene mapping, gene cloning, and genome-wide association analysis (GWAS). Because of the extensive diversity in fruit related traits, such as shape, size, rind thickness and color, flesh texture and color, and content of sugar and carotenoids, watermelons have become one of the model crops for fruit-quality research (Zhu et al., 2016).

Chlorophyll and carotenoids are the main pigments affecting watermelon rind coloration. Because of their essential role in harvesting light energy and converting it into chemical energy, chlorophyll is of great importance in photosynthesis (Fromme et al., 2003). Biosynthesis of chlorophyll belongs to a branch of the tetrapyrrole metabolic pathway (Lange and Ghassemian, 2003) and four distinct sections are included in the biosynthesis progress (Masuda and Fujita, 2008). The first section is synthesis of protoporphyrin IX from 5-aminolevulinic acid (ALA), the precursor of chlorophyll (Hotta et al., 1997). In this progress, ALA is condensed to the monopyrrole, porphobilinogen, and four molecules, and then cyclic tetrapyrrole and uroporphyrinogen III would be synthesized (Grimm, 1998). After decarboxylation and oxidation, protoporphyrin IX is formed at the last step of this section. The second section is the insertion of Mg^{2+} into protoporphyrin IX for Chl a biosynthesis, which is called "the Mg branch" (Walker and Weinstein, 1994). At the last step in the second section, chlorophyllide a would be esterified with a long chain polyisoprenol (geranylgeraniol or phytol) to synthesize Chl a (Tamiaki et al., 2007). The third section is the interconversion of Chl a and Chl b known as "Chl cycle." In this cycle, a portion of Chl a is converted into Chl b by the activity of Chlide a oxygenase (CAO) (Rüdiger, 2002). Chl b can also be reversibly converted to Chl a (Sundby et al., 1986). The last section is the degradation of Chl a (Takamiya et al., 2000; Hörtensteiner, 2006).

Carotenoids have 40-carbon isoprenoids that play essential roles in light harvesting and photoprotection in photosynthetic organisms, and usually provide characteristic colorations of evolutionary adaptive value in plants, fungi, and animals (Britton et al., 2008; Rebeille and Douce, 2011). More than 750 structurally defined carotenoids have been identified in various organisms including bacteria, archaea, fungi, algae, land plants, and animals (Takaichi, 2011). Chloroplasts of green tissues and chromoplasts of flower petals, fruits, and roots are the main sites where carotenoids are synthesized (Yuan et al., 2015). The carotenoids are initially formed by the synthesis of phytoene *via* geranylgeranyl diphosphate (GGPP) through the innermost isoprenoid pathway (Sundby et al., 1986). Then phytoene is further metabolized through desaturations, cyclizations, and hydroxylations to yield various products, such as lycopene, carotenes, and xanthophylls, by a sequence of tandem reactions (Schofield and Paliyath, 2005).

Varying degrees of yellow and green color have been observed in watermelon rind. According to previous studies, the rind color of many plants in the Cucurbitaceae family is controlled by a single gene. A gene for orange fruit in cucumber and another gene

for wax gourd pericarp color were fine mapped, respectively (Li et al., 2013; Jiang et al., 2015). The rind colors of yellow and dark green in watermelon also follow the monogenic inheritance pattern. The gene named *D* for dark green is dominant to the *d* allele for light green rind (Li et al., 2019). One gene named *go* with single recessive inheritance pattern for yellow rind was first reported in 1956 (Barham, 1956). As the fruit matures, color of the fruit rind will change from dark green to golden yellow, and stem and older leaves will become golden yellow (Barham, 1956). Different with gene *go*, another watermelon yellow rind gene following the dominant pattern was mapped to a region on chromosome 4 (Dou et al., 2018). But according to information of the primers sequence and the new released watermelon genome (Watermelon 97103 genome v2) assembled with the PacBio long reads, the dominant gene for yellow rind should be within a region of 729.05 kb but not 59.8 kb (http://cucurbitgenomics.org/organism/21) on the top end of chromosome 4 (Dou et al., 2018).

Recently, watermelon with yellow rind has gained increasing popularity among consumers (Dou et al., 2018), whereas the genes regulating yellow rind and their molecular mechanisms are still unknown in watermelon. In the present study, comprehensive transcriptome analysis for DEG (differentially expressed genes) screening and function prediction between the yellow rind inbred line WT4 and the green rind inbred line WM102 was completed. In addition, with genome resequencing of two parental lines and two DNA pools from the F_2 population, the *yellow rind* (*Clyr*) gene was mapped to a candidate region on chromosome 4 with F_2 population plants by BSA-seq and linkage analysis. These results provide new insight into the molecular mechanism of yellow rind formation and aid in elucidating pigment study in watermelon.

MATERIALS AND METHODS

Plant Materials

WM102 is a watermelon inbred line with a dark green rind, which was artificially self-pollinated for at least four generations selected from the Bush Sugar Baby (accession code: Grif15898; provided by USDA-ARS Germplasm Resources Information Network [GRIN] [www.ars-grin.gov]), and the dark green phenotype is stably expressed in this material. WT4 is a yellow rind inbred line, which was used to cross with WM102 to generate F_1 and F_2 populations for inheritance analysis and gene mapping. All plants were grown in the greenhouse at the Maozhuang Research Station of Henan Agricultural University (Maozhuang, Zhengzhou, at approximately 113.59°N, 34.87°E). The yellow and green rind phenotype were visually observed and recorded during fruit maturation when the different appearance could be easily distinguished. Segregation ratios of yellow/green rind in the F_2 population were analyzed with Chi-square tests (χ^2).

Chlorophyll and Total Carotenoid Content Determination

As rind color of WT4 becomes completely yellow and remains unchanged from 14 days after pollination (**Figure 1**), 1 g sample

of fruit rind at 14 DAP was extracted with a mixture of acetone and alcohol (1:1) using a pestle and mortar till residues became colorless. After complete extraction, the absorbance of the extract was read at 663.2, 646.8, and 470 nm on a spectrophotometer (Shimadzu, Kyoto, Japan) and pigment concentrations were calculated according to Lichtenthaler (Harmut, 1987). Each sample was measured with three biological replicates.

DNA and RNA Extraction

Total genomic DNA from young fresh leaves was extracted using the cetyltrimethylammonium bromide (CTAB) method and the concentration was adjusted to 60 ng/ul (Saghai-Maroof et al., 1984). Fresh samples for RNA extraction were randomly collected from the rinds of three injury-free watermelon fruits in WT4 and WM102 at three different developmental stages (0, 7, and 14 DAP) (**Figure 1**). These samples were immediately frozen in liquid nitrogen, delivered rapidly to the laboratory, and stored at −80– until analysis. For qRT-PCR analysis, total RNA was extracted using the EasyPure® Plant RNA Kit (TRANS) as described by the manufacturer and DNA was removed by digestion with RNase-free DNase. The quality of RNA was assessed with a 1% agarose gel and reverse transcribed to cDNA using a Rever Tra Ace-α-First Strand cDNA synthesis kit (Toyobo).

RNA-Seq Library Construction, Sequencing, and Reads Mapping

The extracted RNA samples were sent to the Biomarker Technologies Co. Ltd (Beijing) for cDNA library construction. The RNA concentration was measured by NanoDrop 2000 (Thermo) and the integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). After the assessment, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and eluted in NEB Next First Strand Synthesis Reaction Buffer. After mRNA was fragmented in small pieces under elevated temperature, first strand cDNA was synthesized with a random hexamer primer and M-MuLV Reverse Transcriptase, and the second strand cDNA synthesis was subsequently completed using DNA Polymerase I and RNase H. Remaining overhangs were then converted into blunt ends. The library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA) to select cDNA fragments of preferentially 240 bp in length. Finally, PCR was performed with Phusion High-Fidelity DNA polymerase. Universal PCR primers, Index (X) Primer, and PCR products were purified by the fdAMPure XP system and library quality was assessed with the Agilent Bioanalyzer 2100 system.

FIGURE 1 | Fruit of watermelon cultivars WT4 and WM102 at critical rind color formation stages. WT4 fruit: 0 DAP **(A)**, 7 DAP **(B)**, and 14 DAP **(C)**. WM102 fruit: 0 DAP **(D)**, 7 DAP **(E)**, and 14 DAP **(F)**.

After the index-coded samples were clustered on a cBot Cluster Generation System using a TruSeq PE Cluster Kit v4-cBot-HS (Illumia), the library preparations were sequenced on an Illumina HiSeq2000 platform and paired-end reads were generated. The RNA sequences have been deposited in the National Center for Biotechnology Information (NCBI) with the accession number of PRJNA549842. Raw reads were filtered by removing low-quality reads and reads containing the adapter, ploy-N. At the same time, Q20, Q30, GC-content, and sequence duplication level of the clean data were calculated. All downstream analyses were based on clean data with high quality score (Q phred) $\geq$ 30 (Q30). The clean reads were aligned to watermelon (97103 V1) reference genome sequences released by the Cucurbit Genomics Data Bank (CuGenDB) (http://cucurbitgenomics.org/) using TopHat 2.0.12 (Trapnell et al., 2009). Only reads with a perfect match or mismatches of no more than two bases were further analyzed and annotated based on the reference genome.

DEGs Screening and Functional Annotation

Expression of three biological replicates was calculated using the DESeq R package to quantify the correlation among biological replicates. DEGs were analyzed with the DESeq2 program based upon reads count (Love et al., 2014). The P-value of DEGs between samples was adjusted using the Benjamini & Hochberg method (Benjamini and Hochberg, 2000). Genes with an adjusted P-value $\leq$ 0.05 were recognized as DEGs. Three pair-wise comparisons between WT4 (yellow rind) and WM102 (green rind) at three stages (0, 7, and 14 DAP) and comparisons of the WT4 among these three different stages were conducted to identify the genes involved in rind color formation. Gene expression was calculated with well-mapped reads, and the results were normalized to the fragments per kilobase of exon per million mapped fragments (FPKM) with the DESeq2 program (Love et al., 2014). To determine the biological significance of the DEGs, a Gene Ontology (GO) enrichment analysis was implemented using the GOseq R package based Wallenius non-central hyper-geometric distribution (Young et al., 2010). GO terms with a corrected P < 0.05 were considered significantly enriched by DEGs. Similarly, KOBAS software was employed to test the statistical enrichment of DEGs in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000). K-means clustering with 9 times repeated was conducted based on Pearson correlation of gene expression profiles (Walvoort et al., 2010).

Expression Level Validation of DEGs by qRT-PCR

Real-time quantitative PCR (qRT-PCR) was used to verify the expression results of the selected genes. RT-qPCR was performed with ABI SYBR green in a one-step real-time PCR system according to the manufacturer's instructions. The gene β-actin was used as the internal reference gene to normalize Ct values of each reaction. Each reaction was performed in a final volume of 16 μl, containing 8 μl SYBR Green PCR Master Mix (Applied Biosystems), 250 nM of each primer, and 50 ng cDNA template. The thermal cycling conditions were 94– for 10 min, followed by 40 cycles of 94– for 15 s, 55– for 30s, and 60– for 1 min, with fluorescence detection at the end of each cycle. Amplification of a single product per reaction was confirmed by melting curve analysis. All reactions were performed with three biological replicates. Expression of some genes with significant different expression level according to the RNA-seq result and the genes within the mapping region were analyzed. Sequences of primers for qPCR are listed **Table S1**.

BSA-Seq Analysis and Preliminary Mapping of Gene *Clyr*

To screen the candidate genomic region responsible for the yellow rind of WT4, 30 yellow-rind plants, and 30 green-rind plants were selected from the F$_2$ population for bulking. The total genomic DNA for each plant was exacted and quantified using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). Then, two DNA pools were constructed by mixing equal amounts of DNA from 30 yellow-rind (Y-pool) and 30 green-rind plants (G-pool). A 5 ug of sample of DNA from the two bulks and two parental lines were used to construct paired-end sequencing libraries, which were sequenced on an Illumina HiSeqTM 2500 platform.

FastQC was used for cleaning and filtering reads (Andrews, 2010). After low-quality and short reads were filtered out, the remaining high-quality reads of each pool were mapped onto the watermelon reference genome sequence 97103 (ftp://cucurbitgenomics.org/pub/cucurbit/genome/watermelon/97103) by BWA (Li and Durbin, 2009). SNP calling followed GATK Best-Practices (McKenna et al., 2010). First, the MarkDuplicates module was used to mark the duplication alignment. Then the BaseRecalibrator and ApplyBQSR modules were used to detect and correct for patterns of systematic errors in the base quality scores, which act as confidence scores emitted by the sequencer for each base. To ensure the accuracy of SNPs identified by GATK, SAMtools software was also used to detect SNPs. The intersection of SNPs that were detected by both GATK and SAMtools software was designated as final SNPs for further analysis. The obtained SNPs and small indels were noted and predicted using SnpEff software (Cingolani et al., 2012), and only the high-quality SNPs with a minimum sequence read depth of five were used for BSA-seq analysis. The SNP-index is an association analysis method to find the significant differences of genotype frequency between the pools, indicated by Δ(SNP-index), and the detail process was followed as previously (Abe et al., 2012; Takagi et al., 2013). The SNP-index is calculated as follows: SNP-index (Green) = ρx/(ρX + ρx), SNP-index (Yellow) = ρx/(ρX + ρx), ΔSNP-index = SNP-index (Yellow) − SNP-index (Green). The Green and Yellow represented the green rind bulk and the yellow rind bulk of the filial generation, respectively. ρX and ρx indicate the number of reads of the alleles in the yellow rind and the green rind parent lines appearing in their pools, respectively. The

difference in each locus between the two pools can be observed through the ΔSNP-index. With respect to the qualitative character, the correlation threshold is the theoretical ΔSNP-index value of the corresponding population and the correlation threshold of the F$_2$ population is 0.67. The regions over the threshold were considered as the associated candidate regions.

All identified SNPs shared across the bulk were considered polymorphic in association studies and two methods were used to identify the candidate regions associated with yellow rind in watermelon: a Euclidean Distance (ED) algorithm and SNP-Index analysis. The calculation of ED was completed using MMAPPR (Mutation Mapping Analysis Pipeline for Pooled RNA-seq) (Hill et al., 2013) and the high ED value suggested that the SNPs in the genomic regions were closely associated with the targeted genes. Δ (SNP-index) was also used to calculate the association at each SNP position between Y-bulk and G-bulk, and previous detailed processes were followed (Abe et al., 2012; Takagi et al., 2013).

To validate the BSA-seq results and further map the target gene, 30 pairs of SSR markers in the candidate region were selected from a genome-wide SSR development (Zhu et al., 2016) and 30 pairs of Indel primers to screen for polymorphism between WT4 and WM102 (**Table S1**). The 30 pairs of Indel primers between WT4 and WM102 were developed using the newly released watermelon genome (ftp://cucurbitgenomics.org/pub/cucurbit/genome/watermelon/97103/v2/) as reference genome. The markers with good polymorphism were further used to genotype an F$_2$ mapping population containing 1,106 plants. PCR amplification of molecular markers and gel electrophoresis were conducted as described in Zhu et al. (2016).

RESULTS

Quantification of Chlorophyll and Carotenoids in the Fruit Rind of Two Parental Lines

To investigate the difference between chlorophyll and carotenoid content in WT4 and WM102, we measured the content of chlorophyll and carotenoid in the fruit rind at 14 DAPs. The content of chlorophyll a and chlorophyll b was significantly reduced in the yellow rind line WT4, which was detected at a very low level, whereas the carotenoids were dramatically increased in WT4 at the same developmental stage (**Figure 2**). The chlorophyll in the green rind line WM102 was at a much higher level compared with that of WT4, and the carotenoids were almost undetectable. This indicated the formation of yellow rind in WT4 was probably because of the reduced chlorophyll and increased carotenoids.

RNA-Seq and Transcript Assembly Identify Novel Genes

To investigate the transcriptomic difference between WT4 and WM102, a total of 18 cDNA libraries were constructed and sequenced for three developmental stages during rind color
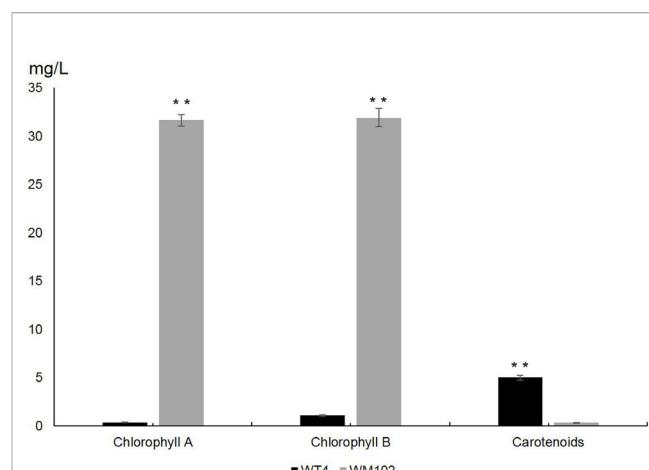
FIGURE 2 | Chlorophylls and carotenoids content in WT4 and WM102 at 14 DAP. Asterisks indicate that the pigment content in WT4 plants was significantly different from that in WM102 plants (**P < 0.01).

formation at 0, 7, and 14 DAP for two parental lines. Approximately 124.71 Gb clean reads were obtained for the 18 cDNA libraries, ranging from 5.96 to 7.98 Gb reads per library. All the clean reads were deposited in the NCBI Short Read Archive (SRA) database under number PRJNA549842.

The clean reads of each sample were separately mapped to the watermelon reference genome 97103 with a mapping rate ranging from 78.31 to 95.83% (**Table S2**). Compared with 23,440 predicted genes in the previous annotations of the watermelon genome, a total of 24,805 genes were identified in the assembly of 18 transcriptomes including 1,365 novel isoforms of unknown genes detected in our study (**Table S3**). The 24,805 genes were used as reference transcripts to determine the read count with HTSeq (Anders et al., 2015). The 1,365 novel genes were further functionally annotated by aligning the sequence to the NCBI non-redundant (Nr) (Pruitt et al., 2005), SwissProt (Boeckmann et al., 2003), GO (Harris et al., 2004), Pfam (Finn et al., 2013), and KEGG (Kanehisa and Goto, 2000) protein databases (e-value <1e-5) by BLAST software. As a result, 818, 429, 433, 387, and 217 genes were successfully annotated in the five above protein databases, respectively (**Table S4**). In addition to the genes whose function is unknown, most of the novel genes were related to "Replication, recombination, and repair," "Transcription," "Translation, ribosomal structure, and biogenesis," "Defense mechanisms," and "Extracellular structures" (**Figure 3**).

The correlation coefficients among three replicates of each period ranged from 0.91 to 0.99, except the coefficients in group 2 (**Table S5**), indicating that most of the three replicates were consistent. The PCA (principal component analysis) analysis indicated that most of the variation in gene expression among different plants was a consequence of the developmental stage. Furthermore, six distinct groups formed within each group, indicated that the transcriptomes of the yellow-rind and green-rind clearly differed from each other (**Figure 4**).

FIGURE 3 | eggNOG Function Classification of novel genes.

DEGs Analysis of WT4 and WM102 During Fruit Rind Coloring

To screen the genes affecting rind color formation, a stringent value of FDR ≤ 0.05 and an absolute value of log2 Ratio ≥ 2 was used as the thresholds to identify the DEGs between WT4 and WM102 at the three stages (0, 7, and 14 DAP). The DEGs between two close stages during fruit rind coloring of WT4 (the earlier stage was considered the control sample, and the later stage was the treated sample). At 0 DAP, the color of WT4 is green-yellow, whereas it is green in WM102. Correspondingly, 581 DEGs between WT4 and WM102 were obtained. Among the 581 DEGs, 302 genes were up-regulated and 279 genes were down-regulated in WT4 (**Figure 5A**). At 7 DAP, the color of WT4 was yellow but it was still green in WM102. Correspondingly, the number of DEGs between WT4 and WM102 was 396, 142 genes, and 254 genes that were up-regulated and down-regulated in WT4 (**Figure 5B**). At 14 DAP, the color of WT4 changed to golden yellow and it is dark green in WM102. A total of 1,385 DEGs were obtained through screening. Among the 1,385 DEGs, 873 were up-regulated and 512 were down-regulated (**Figure 5C**). Because the color of WT4 obviously changed at the three stages, the gene expression level of WT4 at different stages was analyzed. Compared with the genes at 0 DAP, 4,010 DEGs were obtained at 7 DAP in WT4, 2,069 were up-regulated and 1,941 were down-regulated (**Figure 5D**). At 14 DAP, 5,760 DEGs were

compared with the genes at 7 DAP, 3,063 were up-regulated and 2,697 were down-regulated (**Figure 5E**). The number of DEGs of WT4 at different stages is much more than that of WT4 and WM102 at the same periods.

In addition, only 6 DEGs were shared by the above five DEGs groups. Approximately 47 DEGs were shared by the 0-DAP and 7-DAP groups, 142 DEGs were shared by 7-DAP and 14-DAP groups, and 92 DEGs were shared by 0-DAP and 14-DAP groups (**Figure 6**). This further suggested that a common group of genes was activated or deactivated concerning fruit rind coloring.

Identification of DEGs Expression Patterns

To study the DEG expression patterns, the relative expression level of DEGs in WT4 were analyzed by K-means clustering algorithm (Hartigan and Wong, 1979). Results show that there mainly existed nine DEGs expression patterns (subclusters) in WT4 (**Figure 7**). The most prominent group was subcluster_8, in which 2,375 genes were expressed with slight increase levels from stage 1 to stage 3 (**Figure 7H**). In subcluster_6, 1,465 genes were expressed with a slight higher level at 7 DAP than that at both 0 DAP and 14 DAP (**Figure 7F**). In subcluster_7, 969 genes were significantly expression-upregulated at 14 DAP compared with 7 DAP, but with no obvious expression level changes at 7 DAP compared with 0 DAP (**Figure 7G**). A similar pattern was observed for subcluster_4, where genes revealed a significant higher expression level at 14 DAP compared with 0 DAP but a
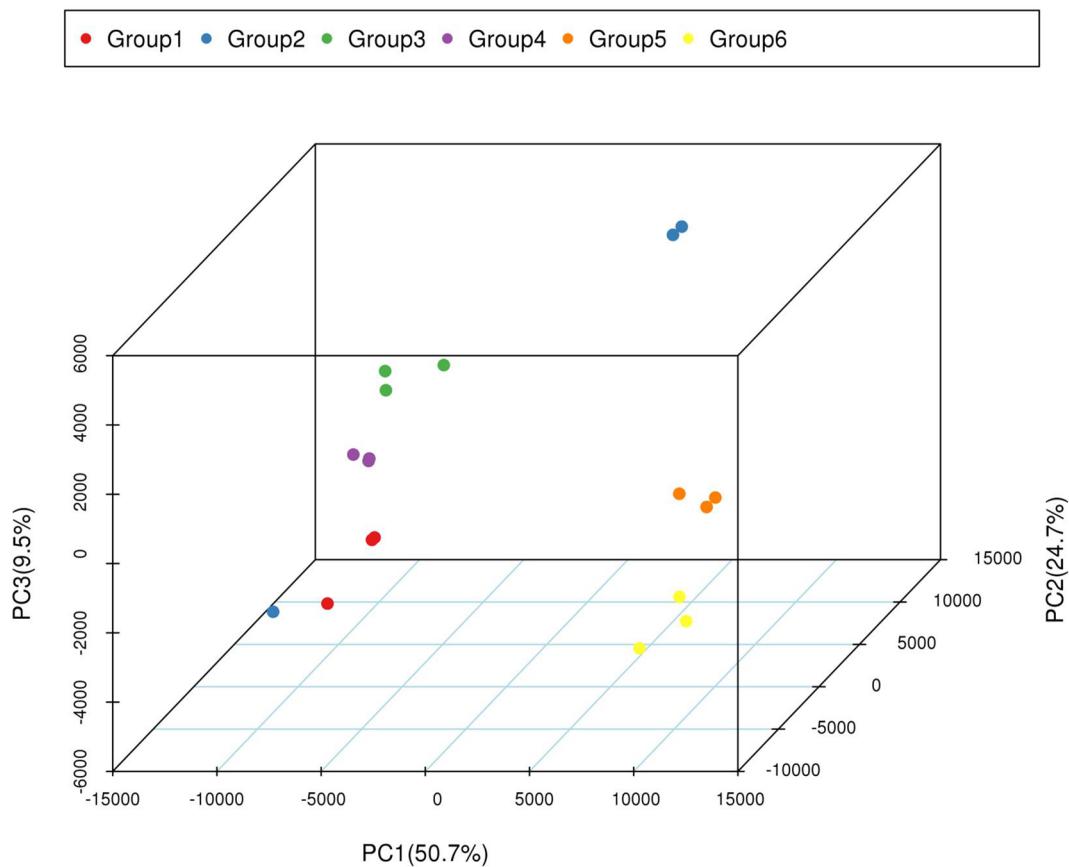
FIGURE 4 | PCA of transcriptomic dataset. Since each sample was repeated three times, Group 1, 3, and 5 contains the repeated samples of WT4 at 0 DAP, 7 DAP, and 14 DAP; Group 2, 3, and 6 contains the repeated samples of WM102 at 0 DAP, 7 DAP, and 14 DAP.

light higher expression level at 7 DAP compared with 0 DAP (**Figure 7D**). Different from subcluster_4 and subcluster_7, 622 genes in subcluster_5 were significantly up-regulated at 7 DAP and 14 DAP compared with expression level at 0 DAP (**Figure 7E**). Contrary to subcluster_5, 671 genes in subcluster_3 were down-regulated at 7 and 14 DAP (**Figure 7C**). A similar expression pattern was exhibited in subcluster_9, but the expression level recovered to a light higher level at 14 DAP (**Figure 7I**). Genes in subcluster_1 were down-regulated at 14 DAP, but the expression level at 7 DAP was similar to that at 0 DAP (**Figure 7A**). Similarly, in subcluster_2, 2,207 genes were significantly down-regulated at 14 DAP, the expression level at 7 DAP was slightly lower than that at 0 DAP, but it was higher than that at 14 DAP (**Figure 7B**). These dynamic gene expression patterns further suggest that yellow rind color is formed *via* a highly complex process.

GO Term Enrichment Analysis of DEGs

To further characterize the function of the DEGs, GO enrichment analysis was completed with GOseq. The top 10 enrichment terms in biological process, cellular component, and molecular function of 0, 7, and 14 DAP were selected as the main

nodes of the directed acyclic graph, respectively (**Table S6**). Of all these enrichment terms, some regarding the metabolism and function of chlorophyll and carotenoids were identified as significant ones. At 0 DAP, the chloroplast membrane was identified as one of the important terms in cellular components, and the gene *Cla011368* (*Cla97C01G001920*) belonging to chloroplast membrane term participates the chlorophyll biosynthetic process and protochlorophyllide reductase activity according to the watermelon genome (**Table S6**). At 7 DAP, the important term chloroplast photosystem II was identified among the enrichment terms in biological process, cellular component, and molecular function (**Table S6**). The chloroplast photosystem II is composed of an inner complex, which contains five chlorophyll a molecules, which have an inner antenna function (Bassi and Dainese, 1992). Genes *Cla001715* (*Cla97C05G096030*) and *Cla021166* (*Cla97C05G081100*) in this term contains the photosystem II oxygen domain and belong to the oxygen evolving enhancer protein 3 family. Except the two important terms above, some other terms, such as photosystem I, plasma membrane, chloroplast inner membrane, light harvesting, photosynthesis, and chlorophyll binding are also recognized (**Table S6**).

FIGURE 5 | Volcano plot showing DEGs between different libraries. P < 0.05 was used as the threshold to judge the significance of difference in gene expression. Red plots represent up-regulated genes; green plots represent down-regulated genes; black plots represent genes with no significant difference. **(A–C)** present genes at 0 DAP, 7 DAP, and 14 DAP between WT4 and WM102; **(D, E)** present genes of WT4 at 7 and 14 DAP compared with 0 and 7 DAP, respectively.

KEGG Pathway Enrichment Analysis of DEGs

A KEGG pathway enrichment analysis of DEGs was conducted to identify the biological pathways of incompatible interaction. The 20 top KEGG pathways with the most representation are shown in **Figure 8**. At 0 DAP, the gene number of plant hormone signal transduction and phenylpropanoid biosynthesis were significantly higher than that of the other pathways (**Figure 8A**). As a major component of plant specialized metabolism, phenylpropanoid biosynthetic pathways provide anthocyanins for pigmentation, flavonoids (Ferrer et al., 2008). The pathways with most genes at 7 DAP are for carbon metabolism, glyoxylate, and dicarboxylate metabolism, photosynthesis, and carbon fixation in photosynthetic organisms (**Figure 8B**). The pathways, such as glyoxylate and dicarboxylate metabolism, photosynthesis, and carbon fixation in photosynthetic organisms were closely related to chlorophyll metabolism. Similar to 0 DAP, the phenylpropanoid biosynthesis pathway, which is concerned with the anthocyanins for pigmentation and flavonoids biosynthesis at 14 DAP contained

many DEGs (**Figure 8C**). Compared with the former stage, the KEGG pathways at 7 DAP and 14 DAP are mainly focused on carbon metabolism and plant hormone signal transduction. The phenylpropanoid biosynthesis pathway was also enriched at 7 DAP (**Figure 8D**). Differently, the pathways, including protein processing in the endoplasmic reticulum and ribosomes also contained many more genes than other pathways (**Figure 8E**).

Because metabolism becomes increasingly active in fruit growing and ripening at 7 DAP and 14 DAP, many DEGs are related with other characters. The chlorophyll a and chlorophyll b content in the WT4 rind was much lower than that of WM102 in the maturation period, whereas the carotenoid content in WT4 was much higher than that of WM102. A total of 56 and 9 DEGs concerning chlorophyll and carotenoid metabolism between WT4 and WM102 at the three stages were obtained (**Figure 9**). Many of these DEGs play important roles in plants chlorophyll and carotenoid biosynthesis. For example, gene *Cla013942* (*Cla97C08G148420*) codes for the Photosystem II Protein.

Photosystem II is the multi-component enzyme of cyanobacteria, algae, and plants that catalyze the light-driven
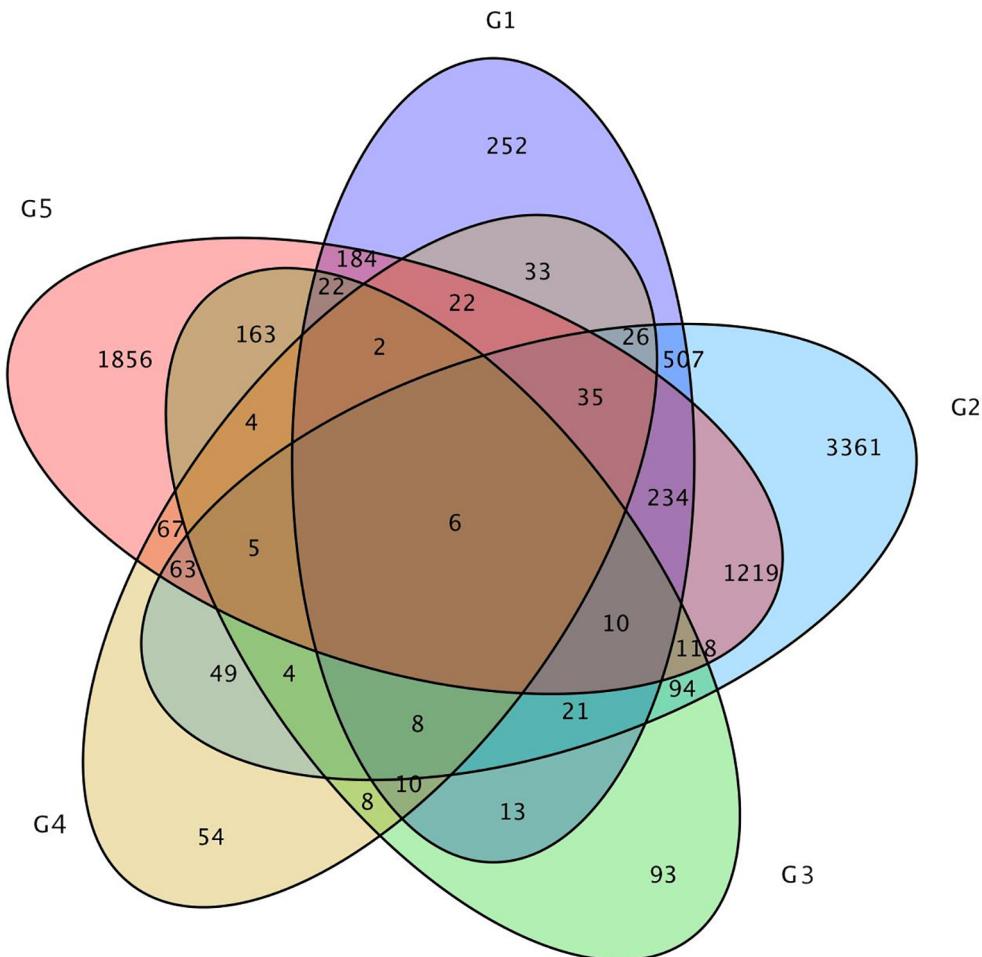
FIGURE 6 | Venn diagram of the relationship between DEG groups. The numbers indicate the DEG number in each DEG group. Group 1, 2, and 3 contains DEGs at 0, 7, 14 DAP between WT4 and WM102, respectively. Group 4 and 5 contains DEGs at 7 and 14 DAP compared with 0 and 7 DAP of WT4, respectively.

oxidation of water to molecular oxygen. This complex consists of more than 20 proteins, including both integral membrane and extrinsically associated proteins (Roose et al., 2007). Gene *Cla008566* (*Cla97C02G049100*) is responsible for coding of magnesium chelatase subunit D. Magnesium chelatase inserts Mg^{2+} into protoporphyrin IX in the chlorophyll and bacteriochlorophyll biosynthetic pathways, which is the key step during chlorophyll a biosynthesis (Willows and Beale, 1998). The chlorophyll and carotenoid mechanisms are very complex, concerning multiple metabolic activities, abnormal expression of each gene in the process may affect pigment biosynthesis and rind color formation.

To validate the RNA-Seq data, qRT-PCR was performed for 13 DEGs identified by RNA-seq analysis. The 13 genes were selected to reflect some of the functional categories and pathways related to chlorophyll and carotenoid biosynthesis. Comparison with the RNA-Seq data showed that the trends in these gene expression patterns were consistent and had a strong positive correlation coefficient ($R^2 = 0.9558$), indicating that the DEGs detected from RNA-Seq analysis were reliable (**Figure 10**).

Mapping of *Clyr* Gene by BSA-Seq Analysis and Linkage Analysis

The yellow rind phenotype in the F_2 population could be easily identified after fruit maturation, and a total of 1,106 plants of the F_2 population were investigated in terms of rind color. Among them, 818 plants were observed with yellow rind, and 288 of them were green rind plants. This was consistent with a 3 to 1 segregation ratio (P = 0.583 in a χ^2 test against 3:1). These results suggested that the *yellow rind* (*Clyr*) in watermelon was controlled by a dominant gene. To explore further the candidate gene regulating yellow rind formation in WT4, a BSA-seq strategy was used to identify the candidate region harboring the *Clyr* gene. We randomly selected 30 yellow rind plants and 30 green rind plants from the F_2 population to mix the Y-bulk and G-bulk for NGS sequencing. After filtering low-quality reads, the resequencing of the two parental lines generated 79,907,647 and 73,483,217 clean reads with 23.94 and 22.02 Gb for WT4 and WM102, respectively. For the two bulks, a total of 30.97 Gb clean data were obtained (15.43 Gb for the Y-bulk and 15.43 Gb for the G-bulk) with an average depth of
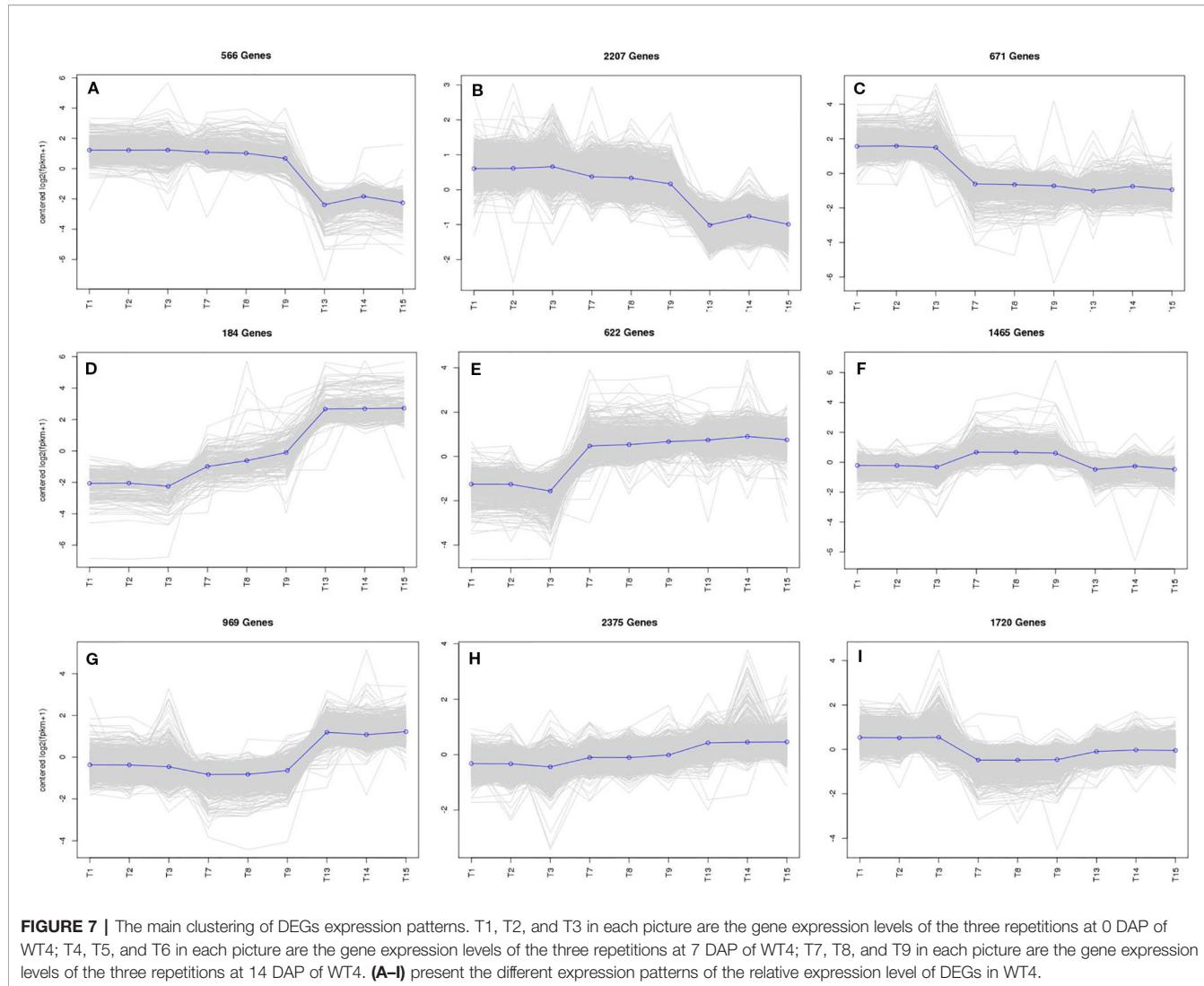
FIGURE 7 | The main clustering of DEGs expression patterns. T1, T2, and T3 in each picture are the gene expression levels of the three repetitions at 0 DAP of WT4; T4, T5, and T6 in each picture are the gene expression levels of the three repetitions at 7 DAP of WT4; T7, T8, and T9 in each picture are the gene expression levels of the three repetitions at 14 DAP of WT4. **(A–I)** present the different expression patterns of the relative expression level of DEGs in WT4.

29 × the genome assembly. There were 111,074 high-quality SNPs detected after filtering SNPs with low coverage and discrepancy between parental lines and bulks. To identify the genomic region associated with yellow rind phenotype, we used the ED algorithm and the SNP-Index to measure the allele segregation of SNPs between the two bulks. In the SNP-index analysis, there was no significant region identified associated with the yellow rind trait. However, there was an obvious peak under the significant threshold which was located in the same candidate region detected by ED analysis. The most significant region associated with yellow rind detected by ED algorithm was on watermelon chromosome 4 from 0 to 8.83 Mb (**Figure 11A**), and the candidate region detected by the SNP-Index was from 4.63 to 7.77 Mb of chromosome 4 (**Figure 11B**).

To validate the BSA-seq results, 30 SSR markers and 30 indel markers in this region on chromosome 4 were selected for polymorphism screening between the two parental lines, WT4 and WM102. Six markers showed clear bands and good polymorphism, and they were further used for genotyping the F_2 segregating population containing 1,106 plants. As a result,

the *Clyr* gene was mapped between primer Indel590358 and the terminal of chromosome 4, covering a physical distance of 91.42 Kb (**Figure 11C**). According to the watermelon reference genome information, a total of three genes exist in the mapped region (**Figure 11C**). But expression analysis shows that expression level of the three genes are not obviously different between WT4 and WM102 (**Figure 12**).

DISCUSSION

Color is a focused trait of consumers for fruits and vegetables. Watermelon with yellow rind has become increasingly desirable for its delightful appearance and high carotenoid contents (Dou et al., 2018). Variation in rind color, including black, dark green, light green, and yellow are exhibited in watermelon (Guo et al., 2013). Chlorophylls and carotenoids are the main pigments influencing the color appearance of plants. According to the different content of these two pigments, the color of plants can vary from dark-green to
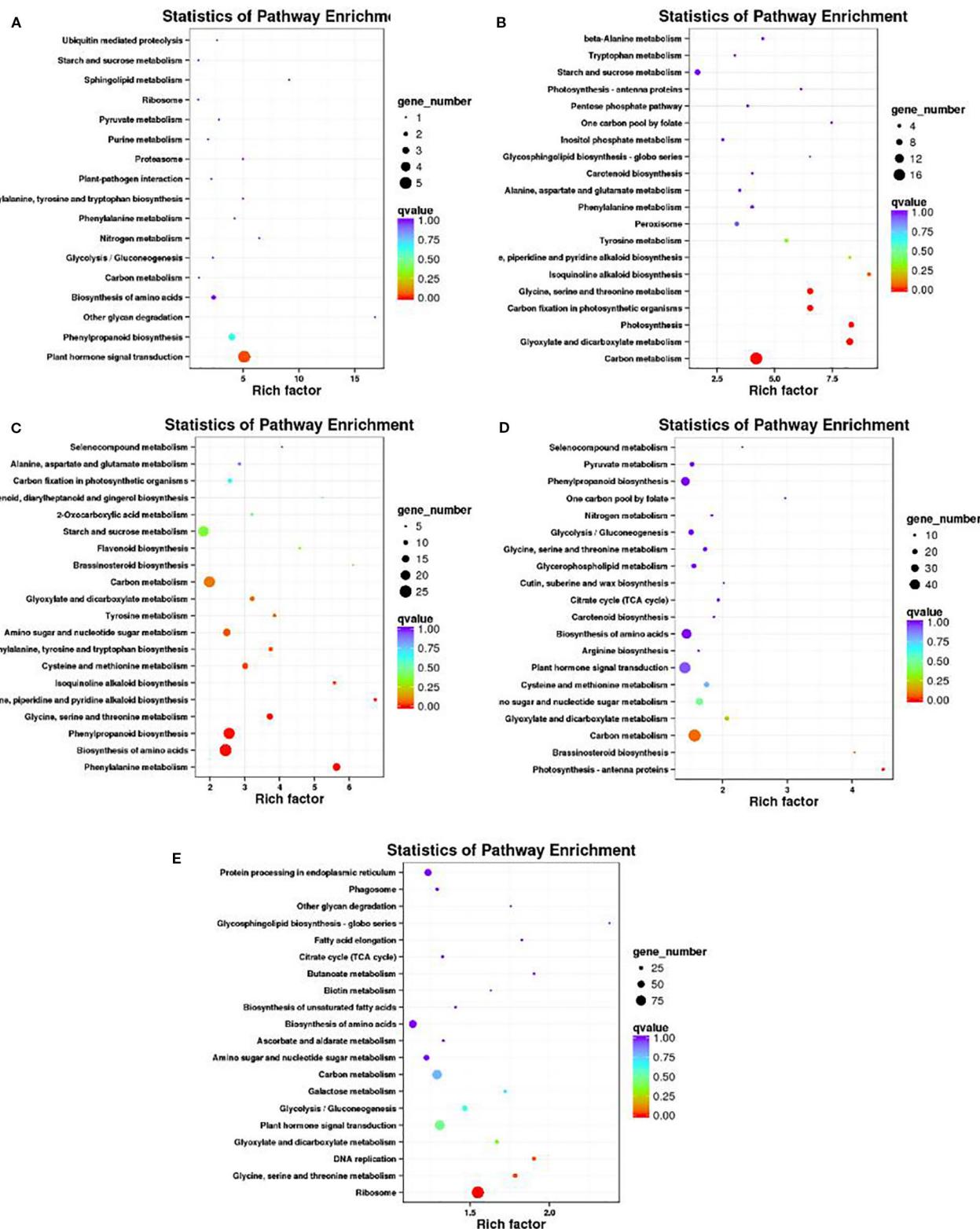
FIGURE 8 | Scatter plot of the KEGG pathway enrichment of DEGs. Rich factor is the ratio of the DEG number to the background number in a certain pathway. The size of the dots represents the number of genes, and the color of the dots represents the range of the q-value. **(A–C)** present the significant terms at 0 DAP, 7 DAP, and 14 DAP between WT4 and WM102; **(D, E)** present the significant terms of WT4 at 7 and 14 DAP compared with 0 and 7 DAP, respectively.
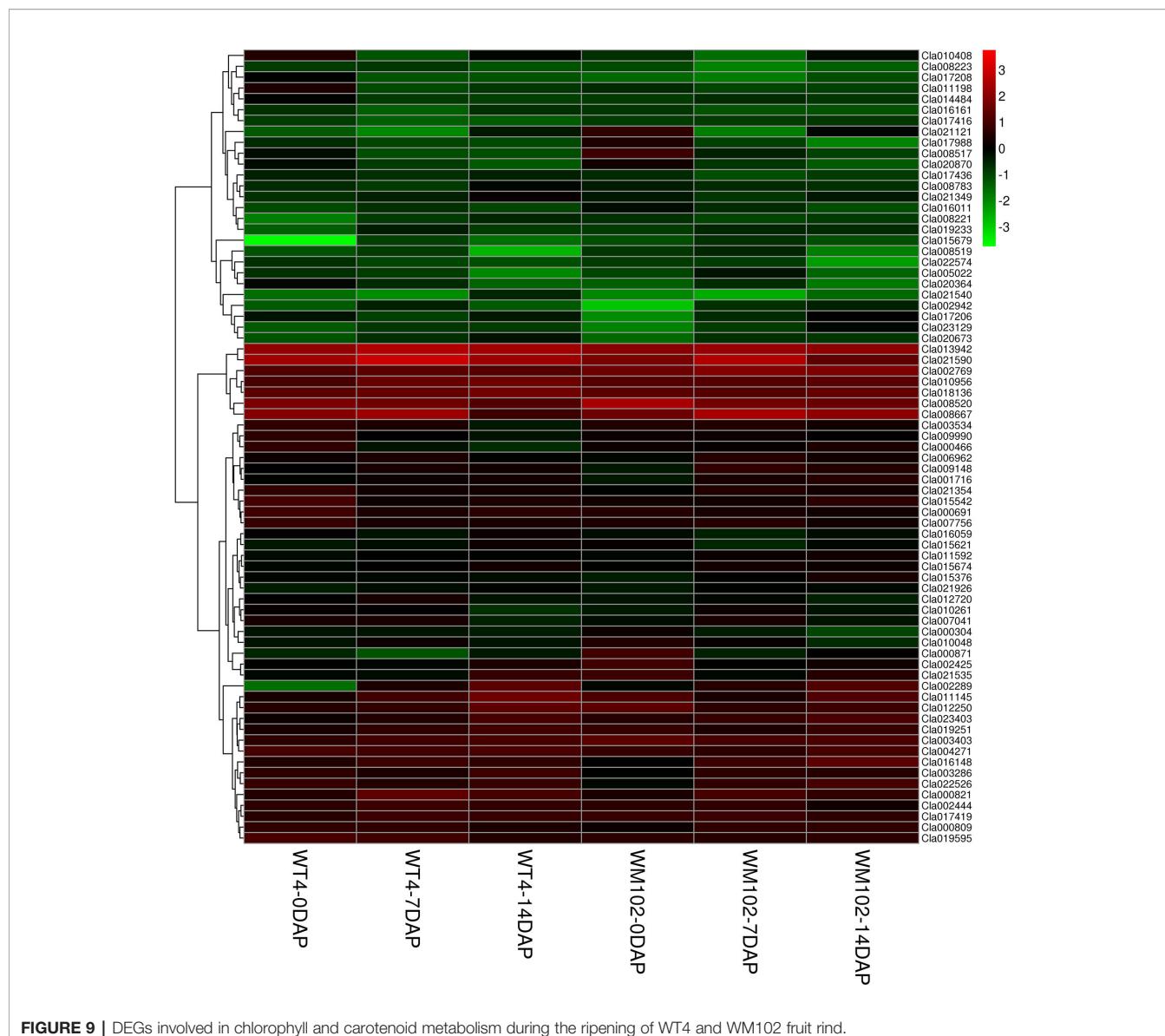
FIGURE 9 | DEGs involved in chlorophyll and carotenoid metabolism during the ripening of WT4 and WM102 fruit rind.
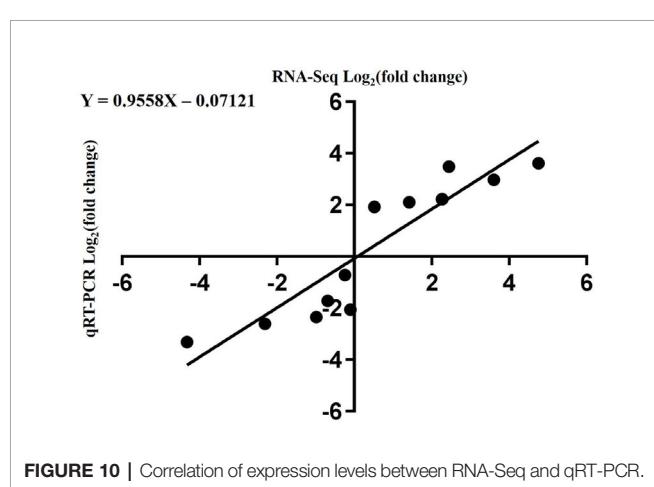
FIGURE 10 | Correlation of expression levels between RNA-Seq and qRT-PCR.

yellow (Jabeen et al., 2017). In WT4, we found a dramatic reduction in chlorophyll a and chlorophyll b, but an increase in carotenoids, implying a close relationship between the yellow appearance and pigment content changes.

The key roles of metabolic pathways during rind color formation were studied with RNA-Seq technology to explore the transcriptomic differences between the two contrasting cultivated watermelon genotypes. A total of 581, 396, and 1,385 DEGs were obtained at 0, 7, 14 DAP. Because color of WT4 changes obvious for the three stages, the gene expression level of WT4 at 7 DAP and 14 DAP were also analyzed. Compared with the genes at 0 DAP, 4,010 DEGs were obtained at 7 DAP and 5,760 DEGs at 14 DAP were obtained compared with the genes at 7 DAP. The number of DEGs for WT4 at different stages was much greater than that of WT4 and WM102 in the same periods, implying a more active metabolism in the later stages and a group of genes that contribute to the
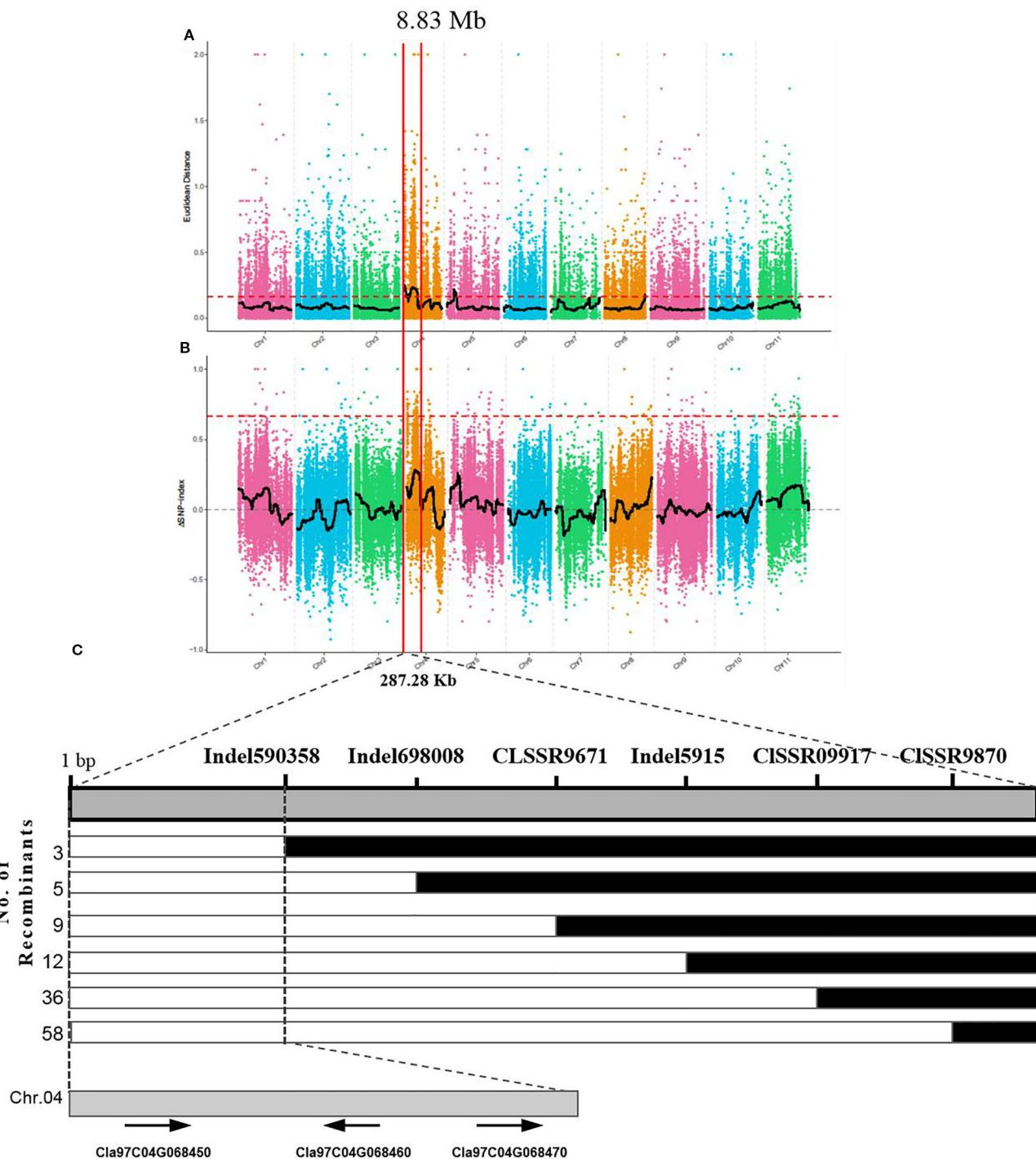
FIGURE 11 | BSA-seq analysis and preliminary mapping of the candidate regions by linkage analysis. BSA-seq analysis using Euclidean Distance (ED) algorithm **(A)** and SNP-Index method **(B)**. The candidate gene was confined to a region of 228.7 kb on the top end of chromosome 4 **(C)**. The colored dots represented the calculated SNP-index (or ΔSNP-index) value, and the black line is the fitted SNP-index (or ΔSNP-index) value. The red dashed line represents the significant threshold.

development and color formation of these two watermelon cultivars. There were different DEG expression patterns in WT4. Some genes were up-regulated during the experimental stages, but some were down-regulated, and different patterns were also exhibited, suggesting a highly complex process concerning yellow rind color formation. Functional enrichment analysis of the DEGs was conducted to identify the most important pathways involved in rind color formation. In addition to the extensively enriched
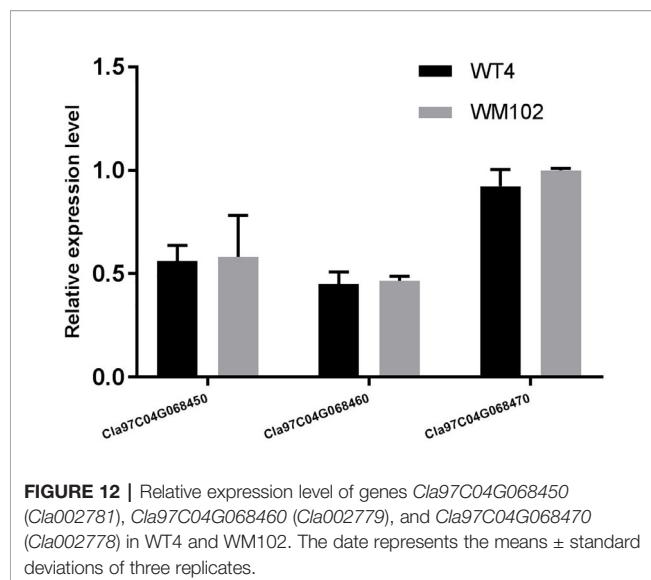
FIGURE 12 | Relative expression level of genes *Cla97C04G068450* (*Cla002781*), *Cla97C04G068460* (*Cla002779*), and *Cla97C04G068470* (*Cla002778*) in WT4 and WM102. The date represents the means ± standard deviations of three replicates.

pathways in WT4 and WM102, some DEGs were found to be involved in chloroplast membrane, plant hormone signal transduction, photosynthesis, carbon fixation in photosynthetic organisms, and might have unique functions in pigment mechanism. These metabolic pathways are also important for color formation in apples and *Arabidopsis* (Miura et al., 2010; El-Sharkawy et al., 2015). In rice, the membrane localized short chain dehydrogenase encoded by gene *NYC1* represents a chlorophyll b reductase that is necessary for catalyzing the first step of chlorophyll b degradation. For the reason that chlorophyll b degradation was required for the degradation of light-harvesting complex II and thylakoid grana in leaf senescence, the rice *nyc1* mutant shows the stay-green phenotype (Sato et al., 2009).

Most of the typical color appearance in cucurbitaceous plants is controlled by a single dominant/recessive gene. In a previous research, the gene for watermelon yellow rind was mapped to a 59.8 Kb interval, but no target gene was found (Dou et al., 2018). But according to the newly released watermelon genome and primer information, the interval in the previous research should be 711.96 kb, not 59.8 kb. Bulked-segregant analysis (BSA) is an efficient method for screening markers tightly linked with the target genes for a given phenotype. It has been widely used for gene mapping study, but utilization of BSA methods requires DNA marker development and genotyping, which is time consuming and labor intensive. Next generation sequencing (NGS) technology could provide new ways to accelerate progress in gene mapping and isolation (Trick et al., 2012). With the BSA method, the gene *Clyr* was quickly mapped to the top end of chromosome 4. To further narrow down the mapped region, a large F_2 population that provides more recombinants was analyzed with the polymorphic primers. There are three genes located in the mapped region. Gene *Cla97C04G068450* (*Cla002781*), *Cla97C04G068460* (*Cla002779*), and *Cla97C04G068470* (*Cla002778*) encodes the DNA glycosylase, the ATPase family AAA domain-containing protein, and DNA-binding storekeeper-related protein, respectively. The

magnesium chelatase reaction is one of important step in chlorophyll biosynthesis pathways. The proteins BchI, BchD, and BchH are required to catalyze the insertion of Mg^{2+} into protoporphyrin IX upon ATP hydrolysis during the magnesium chelatase reaction (Willows et al., 1996; Hansson et al., 2013). AAA proteins are Mg^{2+}-dependent ATPases, which usually play essential roles in proteolysis, membrane fusion, cytoskeletal regulation, protein folding, and DNA replication (Neuwald et al., 1999; Vale, 2000; Ogura and Wilkinson, 2001). Suppression of the ER-Localized NgCDC48, a member of the AAA ATPase superfamily, would make the leaves yellowish and inhibits tobacco growth and development (Bae et al., 2009). The above information suggests that gene *Cla97C04G068460* (*Cla002779*) can be the target gene. But the same expression level of *Cla97C04G068460* between WT4 and WM102 implicates that *Cla97C04G068460* may not be the target gene. And the unchanged nucleotide sequence of the cDNA, gDNA, and about 2,000bp bases before gene *Cla97C04G068460* in WT4 according to the resequecing result further confirmed that gene *Cla97C04G068460* would not be the gene for *Cly*r. With GWAS analysis, three candidate genes associated with rind color and rind stripe were found on chromosomes 4, 6, and 8, corresponding to the rind trait loci S (foreground stripe pattern), D (depth of rind color), and Dgo (background rind color) (Park et al., 2016; Guo et al., 2019). Since *Dgo* gene (*Cla97C04G068530/Cla002769*) encodes a magnesium-chelatase subunit H involved in chlorophyll synthesis, which was not mapped in the candidate region, so the relationship between Dgo and yellow-rind trait in WT4 should be further studied.

In recent studies, a couple of transcription factors regulating plant rind color were identified. For example, a few of MYB-type transcription factors have been reported to affect plant pigment development and rind coloration in cucumber, sweet cherry, tomato, apple, and rice (Furukawa et al., 2007; Adato et al., 2009; Ballester et al., 2010; Li et al., 2013; Jin et al., 2016; Lun et al., 2016; Meng et al., 2016). Except the MYB-type transcription factors, many other transcription factors could determine plant coloration. Such as CsMADS6, a MADS transcription factor in sweet orange (*Citrus sinensis*), could modulate carotenoid metabolism by directly regulating carotenogenic genes (Lu et al., 2018). The F-box gene numbered *MELO3C011980* in melon was also speculated to negatively regulates flavonoid accumulation (Feder et al., 2015). In watermelon, a gene numbered *ClCG08G017810* that encodes a 2-phytyl-1,4-beta-naphthoquinone methyltransferase protein was speculated to be associated with formation of dark green rind color (Li et al., 2019). But according to another study, the transcription factor CmAPRR2 was identified as causative for the qualitative difference between dark and light green rind both in melon and watermelon (Oren et al., 2019). The transcription factors often participate pigment development and rind coloration in plant, but according to the watermelon genome information, no transcription factor was found in the mapped region, the gene for watermelon yellow rind may be not a transcription factor.

Non-protein-coding transcripts including small noncoding RNAs and long noncoding RNAs are reported to play pivotal roles in the epigenetic and post-transcriptional regulation of gene expression during growth, development, and stress response in

plants. Xu found that the expression level of lncRNAs was tightly linked to DNA methylation and that endosperm hypomethylation could promote the expression of linked lncRNAs (Xu et al., 2018). In a dominant wax-less cabbage mutant, the target gene (*NWGL*) was confined to a region approximately 99-kb on the end of cabbage chromosome C08, but no DNA variance was found of the candidate gene (*Bol018504*) in this region. However, its reduced expression abundance and altered DNA methylation level was detected, which was speculated to be one of the possible reasons account for the mutant phenotype (Zhu et al., 2019). Considering the similar dominant mutant style and the no nucleotide changed in the target gene of *NWGL* gene, we predict that appearance of the yellow rind character in WT4 may also be the result of methylated modification. As how the regulators play its key roles during watermelon yellow rind formation, much more work needs to be done. Such as to further narrow down the mapped region and the functional study of the genes and the non-protein-coding transcripts in the mapped region.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are located as follows: Genbank accession for the RNA-sequencing dataset of WT4 and WM102 is PRJNA549842. Genbank accession for the re-sequencing dataset of WT4 and WM102 is PRJNA551784. Genbank accession for the BSA analysis dataset of WT4 and WM102 is PRJNA576063.

AUTHOR CONTRIBUTIONS

LY and YY designed the study. MZ, DS, and XJW performed the RNA isolation and qRT-PCR experiments. DL and XCW performed the data analysis. HY and SY participated in the gene mapping and determination of pigment content. LY, HZ, DS, and DL wrote and revised the manuscript. All authors read and approved the final version of this manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00192/full#supplementary-material

TABLE S1 | Sequences of primers for qPCR and gene mapping.

TABLE S2 | Number of clean reads generated from each sample were sequenced and mapped to the 97103 genome.

TABLE S3 | Sequences of the novel isoforms of unknown genes.

TABLE S4 | Annotation of the novel genes to five protein databases.

TABLE S5 | The correlation coefficients among three replicates of each period.

TABLE S6 | Top 10 enrichment terms in biological process, cellular component, and molecular function in each group.

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

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. (2009). Fruit-surface flavonoid accumulation in tomato is controlled by a *SlMYB12*-regulated transcriptional network. *PloS Genet.* 5 (12), e1000777. doi: 10.1371/journal.pgen.1000777

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169. doi: 10.1093/bioinformatics/btu638

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bae, H., Choi, S. M., Yang, S. W., Pai, H. S., and Kim, W. T. (2009). Suppression of the ER-localized AAA ATPase NgCDC48 inhibits tobacco growth and development. *Mol. Cells* 28 (1), 57–65. doi: 10.1007/s10059-009-0101-4

Ballester, A. R., Molthoff, J., de Vos, R., te Lintel Hekkert, B., Orzaez, D., Fernández-Moreno, J., et al. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor *SlMYB12* leads to pink tomato fruit color. *Plant Physiol.* 152 (1), 71–84. doi: 10.1104/pp.109.147322

Barham, W. S. (1956). A study of the Royal Golden watermelon with emphasis on the inheritance of the chlorotic condition characteristic of this variety. *Proc. Amer. Soc. Hort. Sci.* 67, 487–489.

Bassi, R., and Dainese, P. (1992). A supramolecular light-harvesting complex from chloroplast photosystem-II membranes. *Eur. J. Biochem.* 204 (1), 317–326. doi: 10.1111/j.1432-1033.1992.tb16640.x

Benjamini, Y., and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* 25 (1), 60–83. doi: 10.3102/10769986025001060

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E. A., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31 (1), 365–370. doi: 10.1093/nar/gkg095

G. Britton, S. Liaaen-Jensen and H. Pfander (Eds.) (2008). *Carotenoids, vol. 4: natural functions (Vol. 4)* (Basel: Birkhäuser, Verlag: Springer Science & Business Media).

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

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 (9), e0200617. doi: 10.1371/journal.pone.0200617

El-Sharkawy, I., Liang, D., and Xu, K. (2015). Transcriptome analysis of an apple (*Malus × domestica*) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. *J. Exp. Bot.* 66 (22), 7359–7376. doi: 10.1093/jxb/erv433

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 (3), 1714–1726. doi: 10.1104/pp.15.01008

Ferrer, J. L., Austin, M. B., Stewart, C.Jr., and Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Bioch.* 46 (3), 356–370. doi: 10.1016/j.plaphy.2007.12.009

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Fromme, P., Melkozernov, A., Jordan, P., and Krauss, N. (2003). Structure and function of photosystem I: interaction with its soluble electron carriers and external antenna systems. *FEBS Lett.* 555 (1), 40–44. doi: 10.1016/S0014-5793(03)01124-4

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., et al. (2007). The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* 49 (1), 91–102. doi: 10.1111/j.1365-313X.2006.02958.x

Grimm, B. (1998). Novel insights in the control of tetrapyrrole metabolism of higher plants. *Curr. Opin. Plant Biol.* 1 (3), 245–250. doi: 10.1016/S1369-5266(98)80112-X

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 (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

Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. *Annu. Rev. Plant Biol.* 57, 55–77. doi: 10.1146/annurev.arplant.57.032905.105212

Hansson, M., Lundqvist, J., Sirijovski, N., Al-Karadaghi, S., and Subunits, I. M. C. (2013). Magnesium chelatase: the molecular motor of chlorophyll biosynthesis. *Handb. Porphyrin. Sci.* 28, 41–84. doi: 10.1142/9789814407755_0019

Harmut, A. (1987). Chlorophylls and carotenoids: pigments of photosynthetic membranes. *Method Enzymol.* 148, 350–383. doi: 10.1016/0076-6879(87)48036-1

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi: 10.1093/nar/gkh036

Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: a k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Applied Statistics).* 28 (1), 100–108. doi: 10.2307/2346830

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., Su, Y. C., and Yost, H. J. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23 (4), 687–697. doi: 10.1101/gr.146936.112

Hotta, Y., Tanaka, T., Takaoka, H., Takeuchi, Y., and Konnai, M. (1997). Promotive effects of 5-aminolevulinic acid on the yield of several crops. *Plant Growth Regul.* 22 (2), 109–114. doi: 10.1023/A:1005883930727

Jabeen, A., Kiran, T., Subrahmanyam, D., Lakshmi, D., Bhagyanarayana, G., and Krishnaveni, D. (2017). Variations in chlorophyll and carotenoid contents in Tungro infected rice plants. *IBM J. Res. Dev.* 5, 1–7. doi: 10.4172/2311-3278.1000153

Jiang, B., Liu, W., Xie, D., Peng, Q., He, X., 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 (1), 1035. doi: 10.1186/s12864-015-2220-y

Jin, W., Wang, H., Li, M., Wang, J., Yang, Y., Zhang, X., et al. (2016). The R2R3 MYB transcription factor *PavMYB 10.1* involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (P runus avium L.). *Plant Biotechnol. J.* 14 (11), 2120–2133. doi: 10.1111/pbi.12568

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Lange, B. M., and Ghassemian, M. (2003). Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol. Biol.* 51 (6), 925–948. doi: 10.1023/A:1023005504702

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

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. Genet.* 126 (8), 2187–2196. doi: 10.1007/s00122-013-2128-3

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 (10), 2741–2753. doi: 10.1007/s00122-019-03384-3

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 (12), 550. doi: 10.1186/s13059-014-0550-8

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 (4), 2657–2676. doi: 10.1104/pp.17.01830

Lun, Y., Wang, X., Zhang, C., Yang, L., Gao, D., Chen, H., et al. (2016). A *CsYcf54* variant conferring light green coloration in cucumber. *Euphytica* 208 (3), 509–517. doi: 10.1007/s10681-015-1592-z

Masuda, T., and Fujita, Y. (2008). Regulation and evolution of chlorophyll metabolism. *Photoch. Photobio. Sci.* 7 (10), 1131–1149. doi: 10.1039/b807210h

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi: 10.1101/gr.107524.110

Meng, R., Zhang, J., An, L., Zhang, B., Jiang, X., Yang, Y., et al. (2016). Expression profiling of several gene families involved in anthocyanin biosynthesis in apple (*Malus domestica Borkh.*) skin during fruit development. *J. Plant Growth Regul.* 35 (2), 449–464. doi: 10.1007/s00344-015-9552-3

Miura, E., Kato, Y., and Sakamoto, W. (2010). Comparative transcriptome analysis of green/white variegated sectors in *Arabidopsis* yellow variegated2: responses to oxidative and other stresses in white sectors. *J. Exp. Bot.* 61 (9), 2433–2445. doi: 10.1093/jxb/erq075

Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* 9 (1), 27–43. doi: 10.1101/gr.9.1.27

Ogura, T., and Wilkinson, A. J. (2001). AAA+ superfamily ATPases: common structure-diverse function. *Genes Cells* 6 (7), 575–597. doi: 10.1046/j.1365-2443.2001.00447.x

Oren, E., Tzuri, G., Vexler, L., Dafna, A., Meir, A., Saar, U., et al. (2019). Multiallelic *APRR2* gene is associated with fruit pigment accumulation in melon and watermelon. *J. Exp. Bot.* 70, 3781–3794. doi: 10.1093/jxb/erz182

Park, S. W., Kim, K. T., Kang, S. C., and Yang, H. B. (2016). Rapid and practical molecular marker development for rind traits in watermelon. *Hortic. Environ. Biotech.* 57 (4), 385–391. doi: 10.1007/s13580-016-0005-0

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33 (suppl_1), D501–D504. doi: 10.1093/nar/gki025

Rüdiger, W. (2002). Biosynthesis of chlorophyll b and the chlorophyll cycle. *Photosynth. Res.* 74 (2), 187–193. doi: 10.1023/A:1020959610952

Rebeille, F., and Douce, R. (2011). Biosynthesis of vitamins in plants part B. *Adv. Bot. Res.* 59, 2–292.

Roose, J. L., Wegener, K. M., and Pakrasi, H. B. (2007). The extrinsic proteins of photosystem II. *Photosynth. Res.* 92 (3), 369–387. doi: 10.1007/s11120-006-9117-1

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. W. L. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *P Natl. Acad. Sci. U.S.A.* 81 (24), 8014–8018. doi: 10.1073/pnas.81.24.8014

Sato, Y., Morita, R., Katsuma, S., Nishimura, M., Tanaka, A., and Kusaba, M. (2009). Two short-chain dehydrogenase/reductases, *NON-YELLOW COLORING 1* and *NYC1-LIKE*, are required for chlorophyll b and light-harvesting complex II degradation during senescence in rice. *Plant J.* 57 (1), 120–131. doi: 10.1111/j.1365-313X.2008.03670.x

Schofield, A., and Paliyath, G. (2005). Modulation of carotenoid biosynthesis during tomato fruit ripening through phytochrome regulation of phytoene synthase activity. *Plant Physiol. Bioch.* 43 (12), 1052–1060. doi: 10.1016/j.plaphy.2005.10.006

Sundby, C., Melis, A., Mäenpää, P., and Andersson, B. (1986). Temperature-dependent changes in the antenna size of Photosystem II. Reversible conversion of Photosystem IIα to Photosystem IIβ. *BBA-Bioenergetics* 851 (3), 475–483. doi: 10.1016/0005-2728(86)90084-8

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74 (1), 174–183. doi: 10.1111/tpj.12105

Takaichi, S. (2011). Carotenoids in algae: distributions, biosyntheses and functions. *Mar Drugs* 9 (6), 1101–1118. doi: 10.3390/md9061101

Takamiya, K. I., Tsuchiya, T., and Ohta, H. (2000). Degradation pathway (s) of chlorophyll: what has gene cloning revealed? *Trends Plant Sci.* 5 (10), 426–431. doi: 10.1016/S1360-1385(00)01735-0

Tamiaki, H., Shibata, R., and Mizoguchi, T. (2007). The 17-propionate function of (bacterio) chlorophylls: biological implication of their long esterifying chains in photosynthetic systems. *Photochem. Photobiol.* 83 (1), 152–162. doi: 10.1562/2006-02-27-IR-819

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (9), 1105–1111. doi: 10.1093/bioinformatics/btp120

Trick, ,. M., Adamski, ,. N. M., Mugford, ,. S. G., Jiang, ,. C. C., Febrer, ,. M., and Uauy, ,. C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12 (1), 14. doi: 10.1186/1471-2229-12-14

Vale, R. D. (2000). AAA proteins: lords of the ring. *J. Cell Biol.* 150 (1), F13–F20. doi: 10.1083/jcb.150.1.F13

Walker, C. J., and Weinstein, J. D. (1994). The magnesium-insertion step of chlorophyll biosynthesis is a two-stage reaction. *Biochem. J.* 299 (1), 277–284. doi: 10.1042/bj2990277

Walvoort, D. J., Brus, D. J., and De Gruijter, J. J. (2010). An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Comput. Geosci-UK.* 36 (10), 1261–1267. doi: 10.1016/j.cageo.2010.04.005

Willows, R. D., and Beale, S. I. (1998). Heterologous expression of the Rhodobacter capsulatus BchI,-D, and-H genes that encode magnesium chelatase subunits and characterization of the reconstituted enzyme. *J. Biol. Chem.* 273 (51), 34206–34213. doi: 10.1074/jbc.273.51.34206

Willows, R. D., Gibson, L. C., Kanangara, C. G., Hunter, C. N., and Von Wettstein, D. (1996). Three separate proteins constitute the magnesium chelatase of Rhodobacter sphaeroides. *Eur. J. Biochem.* 235 (1-2), 438–443. doi: 10.1111/j.1432-1033.1996.00438.x

Xu, W., Yang, T., Wang, B., Han, B., Zhou, H., Wang, Y., et al. (2018). Differential expression networks and inheritance patterns of long non-coding RNA s in castor bean seeds. *Plant J.* 95 (2), 324–340. doi: 10.1111/tpj.13953

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14. doi: 10.1186/gb-2010-11-2-r14

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

Zhu, H., Song, P., Koo, D. H., Guo, L., Li, Y., Sun, S., et al. (2016). Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genomics* 17 (1), 557. doi: 10.1186/s12864-016-2870-4

Zhu, X., Tai, X., Ren, Y., Chen, J., and Bo, T. (2019). Genome-wide analysis of coding and long non-coding RNAs involved in cuticular wax biosynthesis in cabbage (Brassica oleracea L. var. Capitata). *Int. J. Mol. Sci.* 20, 2820. doi: 10.3390/ijms20112820

Quantitative Trait Loci for Seed Size Variation in Cucurbits – A Review

Yu Guo[1,2], Meiling Gao[1,2]*, Xiaoxue Liang[1], Ming Xu[1], Xiaosong Liu[1], Yanling Zhang[1], Xiujie Liu[3], Jixiu Liu[3], Yue Gao[3], Shuping Qu[4] and Feishi Luan[4]

[1] College of Life Sciences, Agriculture and Forestry, Qiqihar University, Qiqihar, China, [2] Heilongjiang Provincial Key Laboratory of Resistance Gene Engineering and Preservation of Biodiversity in Cold Areas, Qiqihar, China, [3] Qiqihar Horticultural Research Institute, Qiqihar, China, [4] College of Horticulture, Landscape Architecture, Northeast Agricultural University, Harbin, China

Cucurbits (Cucurbitaceae family) include many economically important fruit vegetable crops such as watermelon, pumpkin/squash, cucumber, and melon. Seed size (SS) is an important trait in cucurbits breeding, which is controlled by quantitative trait loci (QTL). Recent advances have deciphered several signaling pathways underlying seed size variation in model plants such as Arabidopsis and rice, but little is known on the genetic basis of SS variation in cucurbits. Here we conducted literature review on seed size QTL identified in watermelon, pumpkin/squash, cucumber and melon, and inferred 14, 9 and 13 consensus SS QTL based on their physical positions in respective draft genomes. Among them, four from watermelon (*ClSS2.2*, *ClSS6.1*, *ClSS6.2,* and *ClSS8.2*), two from cucumber (*CsSS4.1* and *CsSS5.1*), and one from melon (*CmSS11.1*) were major-effect, stable QTL for seed size and weight. Whole genome sequence alignment revealed that these major-effect QTL were located in syntenic regions across different genomes suggesting possible structural and functional conservation of some important genes for seed size control in cucurbit crops. Annotation of genes in the four watermelon consensus SS QTL regions identified genes that are known to play important roles in seed size control including members of the zinc finger protein and the E3 ubiquitin-protein ligase families. The present work highlights the utility of comparative analysis in understanding the genetic basis of seed size variation, which may help future mapping and cloning of seed size QTL in cucurbits.

Keywords: cucurbits, watermelon, pumpkin/squash, cucumber, melon, seed size, QTL, comparative analysis

INTRODUCTION

Seed is the start and end points of plant life, and the important determinants of growth and development. Seed size is a key agronomic characteristic of evolutionary fitness in plants during domestication and breeding in many crops with seeds as the main product organ, and a key factor affecting seed yield, eating quality, tolerances to environmental stresses (Song et al., 2007; Bai et al., 2010; Gu et al., 2017; Smitchger and Weeden, 2018). In general, large seeds have more advantages than small ones in crop production because large seeds have bigger endosperm or cotyledons that can provide more nutrients for seedling establishment (Chen et al., 2006). Large seeds also have faster germination rate than small seeds, resulting in stronger seedlings that can better compete for light and nutrition, and stronger resistance or

tolerance to adverse environmental conditions (Silvertown, 1981; Coomes and Grubb, 2003; Gómez, 2004). On the other hand, small seeds may have an advantage in seed transmission, thus affect the reproductive capacity of offspring (Wu et al., 2006).

For many crops, seed size is an important target of selection during domestication (Xia et al., 2018). The seeds of wild plants of modern crops are usually small and roundish in shape, while the domesticated ones in general are much larger. Changes in seed size are the results of natural selection and artificial selection in adaptation to different environments or human needs during domestication. Seed size varies greatly among crop plants (Stanton, 1984; Gómez, 2004; Moles et al., 2005). For crops with seeds as the final target of production (for example, rice, wheat and soybean), seed size is also the determinant of productivity and yield (Fan et al., 2006; Chettoor et al., 2016; Hu et al., 2018). As such, increase of seed size has been one of the primary goals in breeding in cereal crops (Zhang et al., 2016; Savadi, 2018).

Crops in the Cucurbitaceae family, which often referred to as 'cucurbits,' include several economically important fruit vegetables worldwide such as watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai], melon (*Cucumis melo* L.), cucumber (*Cucumis sativus* L.), and pumpkin/squash (*Cucurbita* spp.). Cucurbits are generally prized for their delicious fruits, and the seeds are also good sources of vegetable oil and proteins. For example, the watermelon seeds are rich in oil (Baboli and Kordi, 2010) and protein (Al-Khalifa, 1996; Baboli and Kordi, 2010). However, seed size and weight within cultivated cucurbits also vary considerably depending on their intended uses, which is especially true in watermelon and pumpkin whose seeds have dual uses and seed size is an important target in breeding (Silvertown, 1981; Xiang et al., 2002; Tian et al., 2012). On the other hand, seeds of melon and cucumber do not seem to be under selection in modern breeding, thus which do not seem show as much diversity in seed size and color as those in watermelon and pumpkin/squash (**Figure 1**). Therefore, the cucurbits may offer a good system to understand the differential roles of artificial selection for seed-related traits during cucurbits breeding.

In cucurbits, seed size and weight are quantitatively inherited in nature, and the genetic basis is still poorly understood. However, in recent years, rapid progress has been made in genome sequencing among cucurbits. So far draft genomes have been developed for cucumber (Huang et al., 2009; Yang L. M. et al., 2012; Li Q. et al., 2019), melon (Garcia-Mas et al., 2012; Ruggieri et al., 2018), watermelon (Guo et al., 2013), bottle gourd, *Lagenaria siceraria* (Wu et al., 2017), bitter gourd, *Momordica charantia* (Urasaki et al., 2017), pumpkin/squash (Sun et al., 2017; Montero-Pau et al., 2018), and wax gourd, *Benincasa hispida* (Thunb.) Cogn. (Xie et al., 2019). These advances have made it possible to understand the genetic architecture of seed size variation in cucurbits through QTL studies and genome-wide association analysis. Thus in this report, we attempt to summarize seed size QTL identified in four major cucurbits (watermelon, pumpkin/squash, cucumber, and melon). From these QTL, we inferred consensus seed size (SS) QTL across different cucurbits. We found structural and functional conservation of consensus SS

QTL, as well as unique QTL in each crop. Through comparative analysis across four major cucurbits, we identified potential candidate genes in syntenic regions among several major-effect SS QTL.

OVERALL VIEW OF THE GENES AND REGULATORY NETWORKS CONTROLLING SEED SIZE VARIATION IN PLANTS

In angiosperms, a mature seed is composed of three parts: the embryo, the endosperm and the seed coat, all of which contribute to seed size. Seed development involves many complex processes and numerous genes (e.g., Chaudhury et al., 2001; Sun et al., 2010; Figueiredo and Kohler, 2014; Li and Li, 2016). Mutations in many of those genes may result in seed size change. Large amount of work has been done to understand the genetic control of seed size in primarily the model plants such as *Arabidopsis thaliana*, and rice (*Oryza sativa* L.). Major genes and regulatory pathways that play important roles in seed size control in plants are summarized in **Supplementary Table S1**. Major pathways include the ubiquitin pathway (Li et al., 2008; Xia et al., 2013; Li N. et al., 2019), the *IKU* pathway (Luo et al., 2005; Zhou et al., 2009; Wang et al., 2010), and hormone signaling pathways (Okushima et al., 2005; Du et al., 2017; Hu et al., 2018; Assefa et al., 2019), as well as other regulators (e.g., Huang et al., 2011; Wang S. et al., 2012; Wang et al., 2015; Si et al., 2016).

Several ubiquitin pathway genes have been shown in control of seed size (Li et al., 2008). The ubiquitination process involves ubiquitin activating enzymes (E1s), ubiquitin conjugating enzymes (E2s), and ubiquitin ligases (E3s) (Li et al., 2018b). In Arabidopsis, *DA1* encodes a ubiquitin receptor and acts as a negative regulator in seed size and organ development (Li et al., 2008; Dong et al., 2017). *DA2* encodes a RING-type E3 ligase with a similar function as *DA1* (Xia et al., 2013). The rice *GRAIN WIDTH AND WEIGHT2* (*GW2*) is a homolog of *DA2*; over-expression of *GW2* and *DA2* restricts seed/organ size (Song et al., 2007; Xia et al., 2013; Dong et al., 2017). The maize *ZmGW2-CHR4* and wheat *TaGW2* are homologs of rice *GW2*, both of which have also been shown to control grain size (Li et al., 2010; Su et al., 2011). In addition, the *Brassica napus* HECT E3 ligase gene *BnaUPL3.C03* regulates seed weight; its down-regulation leads to higher seed weight per pod (Downes et al., 2003; Charlotte et al., 2019).

In Arabidopsis, three genes, *HAIKU1* (*IKU1*), *HAIKU2* (*IKU2*), and *MINISEED3* (*MINI3*) in the same HAIKU (IKU) pathway coordinate to control seed size by influencing endosperm development during early seed growth (Li N. et al., 2019). *IKU1* is a VQ motif-containing protein (Wang et al., 2010). *IKU2* and *MINI3* encode a leucine-rich repeat (LRR) kinase, and a WRKY transcription factor, respectively (Garcia et al., 2003; Luo et al., 2005). The *IKU1*, *IKU2* and *MINI3* mutants all show reduced endosperm size, and thus smaller seeds (Garcia et al., 2003; Luo et al., 2005; Wang et al., 2010). In addition, *SHORT HYPOCOTYL UNDER BIUE1* (*SHB1*) is an upstream regulatory
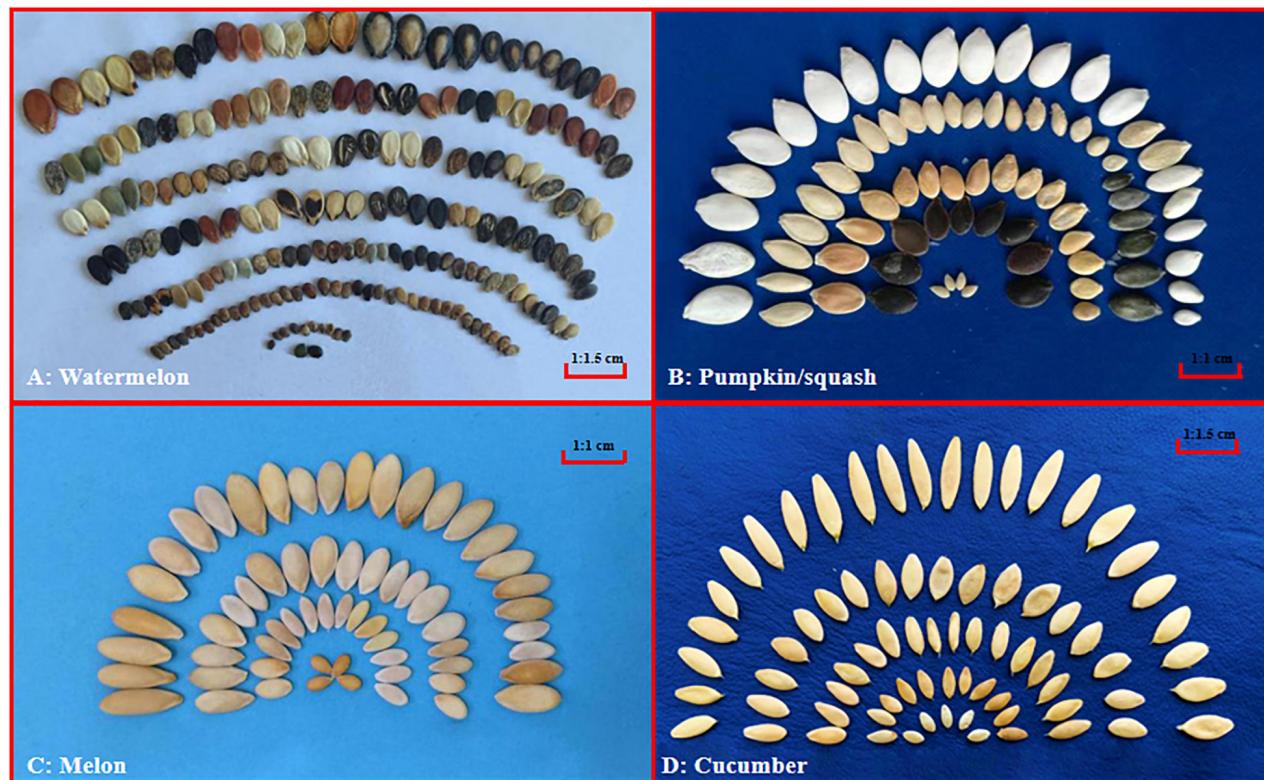
FIGURE 1 | Seed size variation among cucurbit crops. Representative seeds of watermelon (*Citrullus lanatus*), pumpkin/squash (*Cucurbita*), melon (*Cucumis melo*) and cucumber (*Cucumis sativus*) are shown in **(A–D)**, respectively. All images were taken by the authors.

factor affecting the expression of *MINI3* and *IKU2* by interacting with other proteins in the early stage of seed development leading to enlarged seed cavity and improved endosperm growth (Zhou et al., 2009). Over-expression of *IKU2* in canola (*B. napus*) increases seed size and seed yield (Xiao et al., 2016).

Phytohormones, such as brassinosteroids (BRs), gibberellin (GA), auxin and cytokinin (CK) play important roles in seed development which directly control seed size by enhancing cell proliferation, affecting embryo and endosperm development (Jiang et al., 2013; Du et al., 2017; Hu et al., 2018; Assefa et al., 2019; Li Y. D. et al., 2019). Many genes regulate seed development through integration into plant hormone metabolism. For example, *ARF2* (*AUXIN RESPONSE FACTOR 2*) regulates seed size by limiting cell proliferation in the maternal integument (Okushima et al., 2005; Schruff et al., 2005). SRS5 (SMALL AND ROUND SEED), an α-tubulin protein, positively regulates grain length independent of the BR signaling pathway (Segami et al., 2012, 2017). In soybean, over-expression the P450 gene *GmCYP78A5* results in increasing seed weight (Du et al., 2017). The rice gene *qTGW3* encodes OsSK41 that interacts with *OsARF4* (*AUXIN RESPONSE FACTOR 4*); loss-of-function mutant of OsSK41 has increased grain length and weight (Hu et al., 2018). The grape gene *VvHB58* has a potential function in regulating seed number and development by impacting auxin, gibberellin, and ethylene signaling pathways (Li Y. D. et al., 2019). In soybean, *Glyma.02g04660* encodes a histidine phosphotransfer

protein ortholog that controls seed width through the cytokinin-mediated regulating pathway (Assefa et al., 2019). In *B. nupus*, *BnaA9.CYP78A9* positively regulates seed length by affecting auxin metabolism (Shi et al., 2019).

Many transcription factors have been shown to play important roles in regulating seed development. For example, *APETALA2* (*AP2*), a member of the AP2/EREBP transcription factor gene family, regulates cell expansion in the maternal integument of seed development (Jofuku et al., 2005; Zhang et al., 2016). *AINTEGUMENTA* (*ANT*) is another type of *AP2* transcription factor affecting the seed size by regulating the cell division of integument during ovule development (Mizukami and Fischer, 2000). *TRANSPARENT TESTA GLABRA2* (*TTG2*) encodes a WRKY transcription factor that is mainly expressed in the endosperm and integument with a similar function as *AP2*: it may also affect the seed size by directly regulating integument cell elongation (Johnson et al., 2002; Garcia et al., 2005). The MADS-box genes have been found to control seed size in rice. For example, *OsMADS29* regulates development of the integument or seed coat, and *OsMADS87* may affect seed size in endosperm cellularization period (Prasad et al., 2010; Yang X. et al., 2012; Chen et al., 2016). The *WG7* (*WIDE GRAIN 7*) up-regulates *OsMADS1* expression by directly binding to its promoter, and enhancing histone H3K4me3 enrichment in the promoter and ultimately increases grain width in rice (Huang et al., 2019). The *SlPRE2* encodes a bHLH family transcription

factor, which mediates plant response to gibberellin; silencing of *SlPRE2* decreases tomato seed size by restricting cell expansion (Zhu et al., 2019).

Many other genes or QTL not belonging to above-mentioned categories have been cloned that affect seed/grain size. For example, the rice gene *Os02g0192300* for a zinc finger protein seems to control grain weight (Huang et al., 2011). The rice grain width QTL *GW8* encoding OsSPL16 increases grain width by promoting cell proliferation and grain filling (Wang S. et al., 2012). *GLW7* (*GRAIN LENGTH AND WEIGHT ON CHROMOSOME 7*) also encoding a SPL protein (OsSPL13) positively regulates grain size by increasing spikelet hull cell expansion (Si et al., 2016). The rice *GW7* (*GRAIN WIDTH7*) controls grain size and shape through modulating cell proliferation (Wang et al., 2015). The rice *GS3* (*GRAIN SIZE 3*) also could regulate grain size and weight (Li et al., 2016; Zeng et al., 2020). In Arabidopsis, the cyclin gene *CYCB1;4* controls the final seed size by regulating the cell cycles in maternal tissues and zygotic tissues (Ren et al., 2019).

QTL FOR SEED SIZE (SS) VARIATION IN MAJOR CUCURBIT CROPS

There is rich genetic diversity in seed size among major cucurbits (**Figure 1**), which was the result of selection during domestication or diversifying selection in breeding to adapt to crop production for various purposes (Tan et al., 2013; Wang et al., 2014; Li et al., 2018a; Zhang et al., 2018). In recent years, a number of seed size QTL mapping studies have been conducted in watermelon, pumpkin/squash, cucumber and melon QTL, which allows a comparative analysis of the genetic basis of seed size variation across different cucurbits. However, the names of the QTL are confusing in the literature. For convenience, following recommendations for QTL naming by Pan et al. (2020) and Wang et al. (2020), here we developed a QTL nomenclature for naming QTL for seed-related traits, which are presented in **Table 1**. In brief, SL, SWD, ST, and 100SWT stood for the seed length, seed width, seed thickness and 100-seed weight, respectively. The QTL was named according to Pan et al. (2020) which included information for its location and relative order on chromosome (e.g., *sl1.1* presents the first QTL on chromosome 1). In each crop, the same QTL could be identified in different studies with different genetic backgrounds or environments. By examining their physical locations and LOD support intervals, it is possible to establish a consensus SS QTL for these related traits (Pan et al., 2020). Thus, a consensus SS QTL will be defined based on individual SL, SWD, ST, or 100SWT QTL.

Seed Size QTL in Watermelon

Watermelon could be divided into two types based on its uses: use of the edible flesh and use of seeds. For seed-use watermelons, large seeds are preferred because they are rich in oils and proteins, and easy to crack open (Al-Khalifa, 1996; Baboli and Kordi, 2010; Prothro et al., 2012b), while for edible flesh watermelons, small and few (or no) seeds could increase the pulp proportion and

TABLE 1 | Nomenclature of seed size QTL used in the present research.

QTL type	Traits name	Abbreviation	QTL name	Description of QTL name
Single QTL	Seed length	SL	*sl1.1*	First sl QTL on Chromosome 1
	Seed width	SWD	*swd2.2*	Second swd QTL on Chromosome 2
	Seed thickness	ST	*st3.3*	Third st QTL on Chromosome 3
	100 seed weight	100SWT	*100swt4.1*	First 100swt QTL on Chromosome 4
Consensus QTL	Watermelon seed size	ClSS	*ClSS 1.1*	First consensus SS QTL on watermelon Chromosome 1
	Cucumber seed size	CsSS	*CsSS 4.2*	Second consensus SS QTL on cucumber Chromosome 4
	Melon seed size	CmSS	*CmSS 3.1*	First consensus SS QTL on melon Chromosome 3

improve product quality. Thus, seed size is a target of selection in watermelon breeding, but the breeding direction of selection depends on its end use.

There is a wide range of variation in seed size among watermelon collections (e.g., **Figure 1A**). The watermelon seed could be classified into six representative types based on the size: giant, big, medium, small, tiny, and tomato seed (Yongjae et al., 2009). Poole et al. (1941) were pioneers who paid attention to the inheritance of seed size. They found that the length of large and medium-sized seeds was controlled by two genes (*l* and *s*), and that small seeds were dominant to medium seed size. Later, two gene, *Ti* for *tiny seed* and *ts* for *tomato seed* were also reported (Zhang et al., 1994; Tanaka et al., 1995; Zhang, 1996). Zhang and Zhang (2011) found that a pair of major genes and a pair of recessive genes determine seed size, but additional modifiers are possible. Kim et al. (2015) considered watermelon seed size as a quality trait that was controlled by a single dominant gene. Gao et al. (2016) carried out genetic analysis on seed size-related traits, and found that seed length, width and 1000-seed weight were quantitatively inherited that show continuous variation in segregating populations, and they may be controlled by major QTL. The quantitative nature of seed size variation in watermelon was further observed in other studies (e.g., Prothro et al., 2012a; Zhou et al., 2016). Prothro et al. (2012a) were probably the first to conduct QTL analysis for seed size in watermelon, but pre-draft genome studies were sporadic (Meru and McGregor, 2013). Using a recombinant inbred line (RIL) population from KBS (medium seed, cultivated) × NHM (medium seed, cultivated) and an F_2 population from ZWRM (small seed, cultivated) × PI244019 (medium seed, *Citroides*), Prothro et al. (2012a) conducted QTL mapping on seed size related traits and identified 13 QTL on four linkage groups (LG2, LG4, LG9 and LG11). In both populations, three major QTL with phenotypic variance explained (PVE) being from 26.9 to 73.6% for SL, SWD and 100SWT, were identified at the same
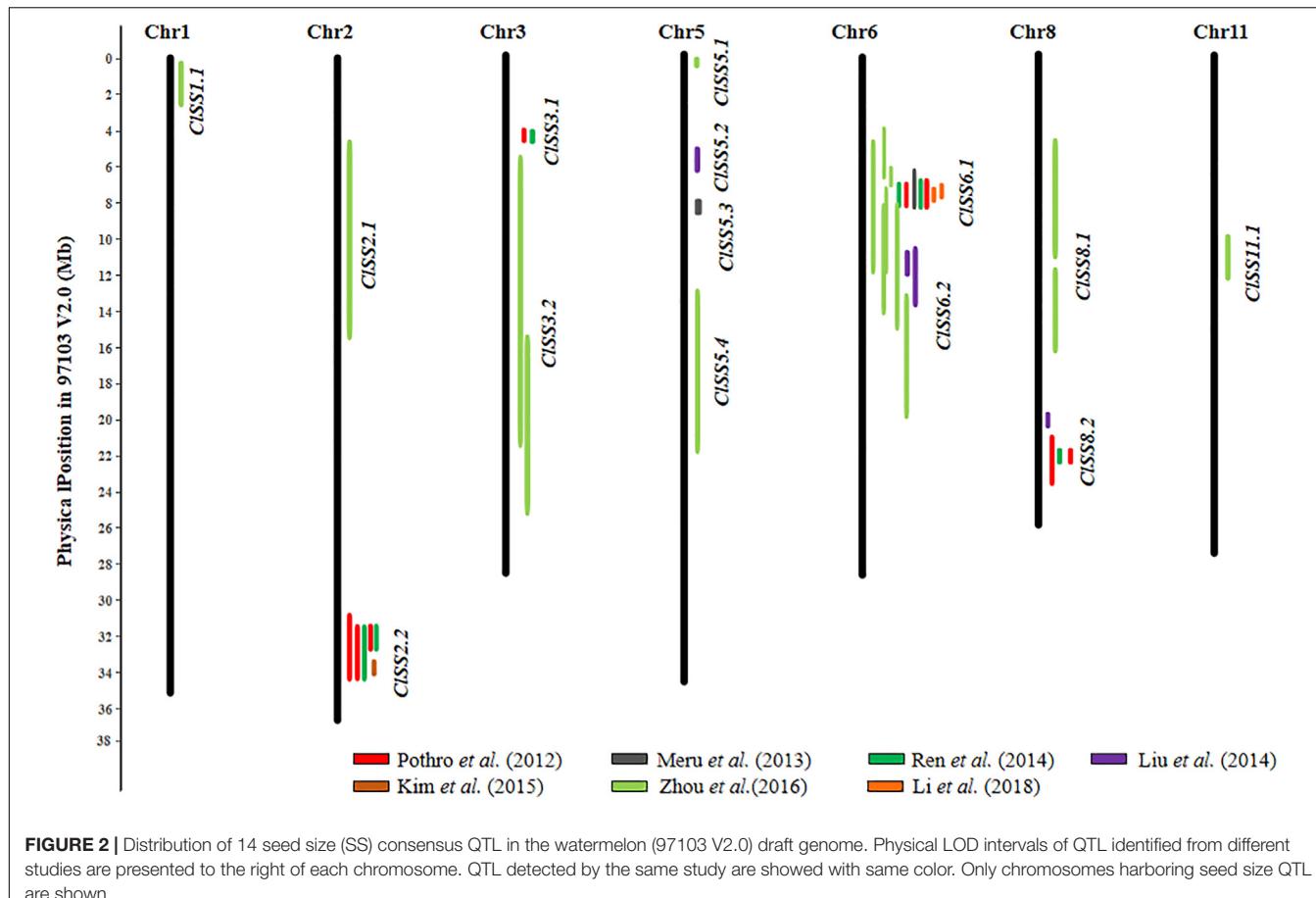
FIGURE 2 | Distribution of 14 seed size (SS) consensus QTL in the watermelon (97103 V2.0) draft genome. Physical LOD intervals of QTL identified from different studies are presented to the right of each chromosome. QTL detected by the same study are showed with same color. Only chromosomes harboring seed size QTL are shown.

region on Chr6 (LG2). In addition, they also detected a major-effect QTL (PVE = 25.6%) for SWD on Chr2 (LG9). Using an F_2 population from the normal type (PI 279261) × egusi type (PI 560023) cross, Meru and McGregor (2013) detected 3 major-effect QTL (PVE = 34.4–60.8%) on Chr6, which seem to be the same chromosomal region as identified by Prothro et al. (2012a).

The release of the watermelon draft genome (Guo et al., 2013) greatly facilitated QTL mapping studies. Ren et al. (2014) verified the major QTL on Chr2 and Chr6 reported by Prothro et al. (2012a). Several studies all detected SS-related QTL in the same Chr6 region (Sandlin et al., 2012; Prothro et al., 2012a,b; Ren et al., 2012). Meanwhile, using an F_2 population developed from the cross between wild egusi type (PI 186490) and cultivated type (LSW-177) watermelons, Liu et al. (2014) identified 1 SL QTL (QSL) and 1 SWD QTL (QSWD1) on Chr6 (PVE = 18.8 and 15.7%), respectively. With the same population, Zhou et al. (2016) mapped more SS QTL (PVE = 3.2–28.9%) on Chr6 including 3 SL, 4 SWD, 1 ST, and 2 100SWT QTL. Kim et al. (2015) detected a 20-seed-weight QTL on Chr2 with an F_2 population from the cross between the open-pollinated cultivar 'Arka Manik' and Jubilee-type inbred line 'TS34.' In our recent study, we identified a major-effect QTL, SS6.1 for seed size on Chr6 using an F_2 population developed from the cross between a small size line and a medium size line. The PVE for QTL of SL, SWD, and 100SWT in this population was as high as 48.5, 42.2, and 45.3%, respectively. Interestingly, in an F_2 population from the cross between two watermelon inbred lines with medium size seeds, we detected a major-effect QTL (PVE = 35.5–50.1%) for SL, SWD, and 20SWT (20 seed weight) on Chr2 (Gao et al. unpublished data).

Li et al. (2018a) conducted fine mapping of QTL for seed-related traits in watermelon, and identified a major-effect QTL qSS6 for seed size. The PVE of the QTL for SL, SWD, and 1000SWT (1000-seed weight) in this population were as high as 94.1, 95.3, and 93.0%, respectively. They further identified a region of qSS6 harboring three candidate genes including Cla009291, Cla009301, and Cla009310. Among them, Cla009291 encodes the MDR protein mdtK, which is differentially expressed at different seed developmental stages between large- and small-seeded lines. Cla009301 is a homolog of SRS3 (SMALL AND ROUND SEED) for a BY-kinesin-like protein 10 that is a seed size regulator in rice (Kanako et al., 2010). Cla009310 encodes an unknown protein that was proposed to be a candidate for qSS6 based on an SNP in the first exon between the two parental lines (Li et al., 2018a).

To summarize, so far 20 SL, 19 SWD, 3 ST, and 19 100SWT QTL (total 61) have been identified in watermelon. The details of these QTL are presented in **Supplementary Table S2**. Additional information for each QTL is also provided including flanking markers with sequences, their chromosomal
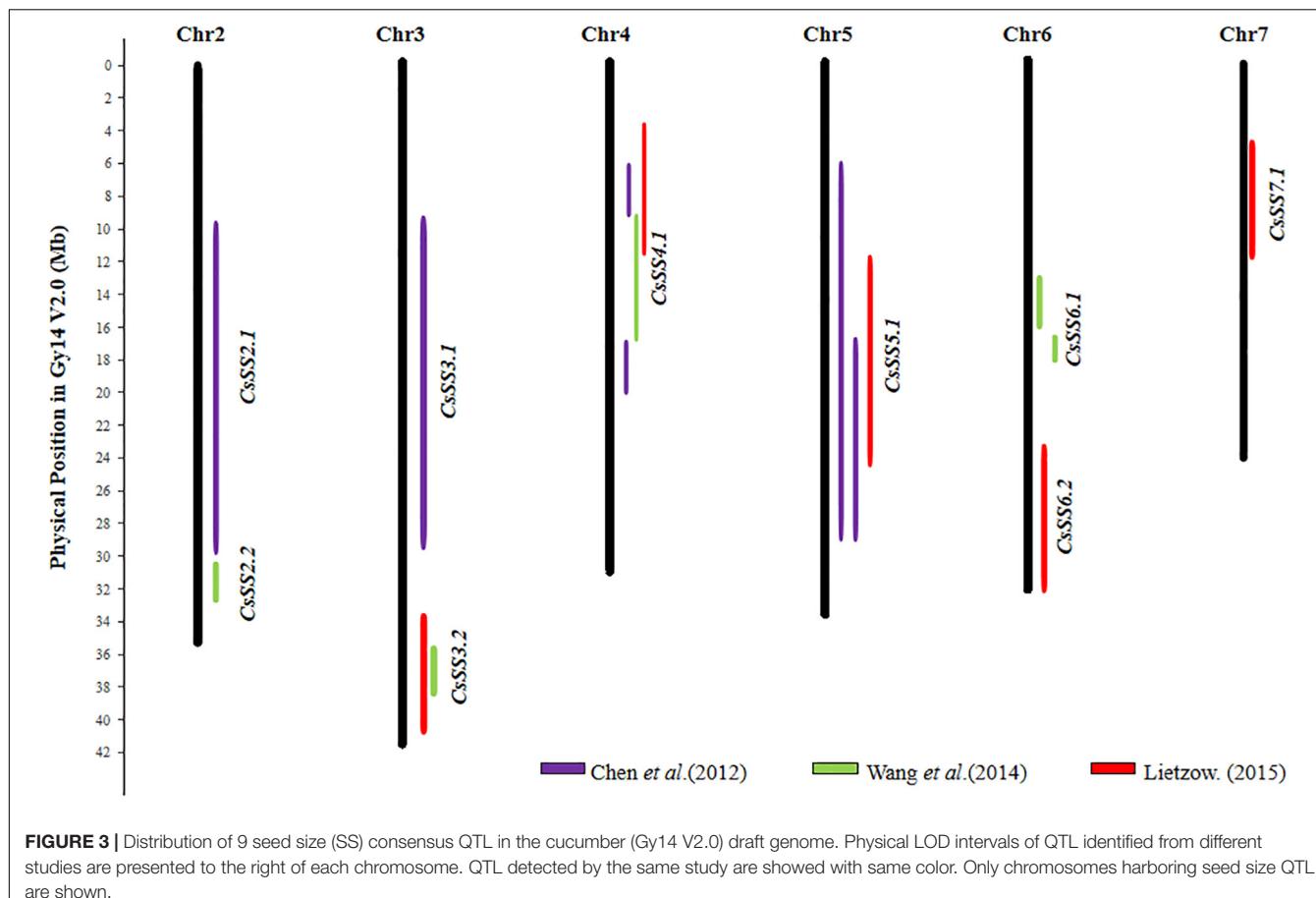
FIGURE 3 | Distribution of 9 seed size (SS) consensus QTL in the cucumber (Gy14 V2.0) draft genome. Physical LOD intervals of QTL identified from different studies are presented to the right of each chromosome. QTL detected by the same study are showed with same color. Only chromosomes harboring seed size QTL are shown.

positions, mapping populations and PVE are presented in **Supplementary Table S2**. Based on their chromosomal locations and LOD support intervals, 14 consensus SS QTL were inferred from different studies. The information of all 14 consensus SS QTL including independent QTL from each study is listed in **Supplementary Table S3**, their chromosomal positions are visually illustrated in **Figure 2**. The 14 consensus SS QTL were distributed on 7 of the 11 watermelon chromosomes. Of them, 4 QTL (*ClSS2.2*, *ClSS6.1*, *ClSS6.2*, and *ClSS8.2*) could be detected in at least two populations/studies. *ClSS2.2* was only detected by the populations constructed from two cultivated lines, but *ClSS6.1* and *ClSS6.2* could be detected in populations derived from cultivated and wild watermelon lines. Especially, *ClSS6.1* and *ClSS6.2* were identified in most studies with F_2 and RILs populations derived from different seed size materials, suggesting that this QTL may play the most important role in seed size/weight determination. Meanwhile, some 'consensus' QTL such as *ClSS1.1* and *ClSS11.1* were only detected in a single population/study (e.g., Zhou et al., 2016), which should be considered putative pending validation in future studies.

Seed Size QTL in Pumpkin/Squash

Pumpkin/squash (*Cucurbit* spp.) are important cucurbit crops cultivated worldwide. The seed size of seed-use pumpkin affects

not only seed yield, but may also impact fruit quality. For example, Paris and Nerson (2003) found that pumpkin SL and SWD were positively correlated with the fruit size, while the seed shape was negatively correlated with the fruit shape. There was larger heterosis of SWD than that of SL (Xu et al., 2006). Seed number per fruit and 100SWT directly affect the yield of seed-use pumpkin (Wang P. et al., 2012). Meru et al. (2018) found that seed size was positively correlated with oil content, but negatively with seed protein in pumpkin. However, little is known about the genetic basis of seed size variation in pumpkin/squash. So far, relatively few QTL have been identified underlying seed size variation in pumpkin/squash. Tan et al. (2013) developed an F_2 population using the Indian large-grain '0515-1' and the small-grain '0460-1-1' pumpkin lines; QTL analysis detected 4 major-effect QTL controlling SWD on LG2, LG3 and LG4 with PVE from 2.9 to 29.7%. In *C. maxima*, Wang et al. (2019) developed a high-density genetic map with the F_2 population from the cross between two squash inbred lines ('2013-12' and '9-6'). In QTL mapping, they identified 10 QTL on six chromosomes (Chr) (Chr4, 5, 6, 8, 17, and 18), including 4 for SL, 4 for SW, and 2 for HSW (100-hundred-seed weight) with the PVE ranging from 7.0 to 38.6%. The major-effect QTL *SL6-1* was located on Chr6, and explained 38.6% of observed phenotypic variance. Another major-effect QTL, *SW6-1* was also detected on Chr6 with the PVE of 28.9%. The details of these QTL are presented in
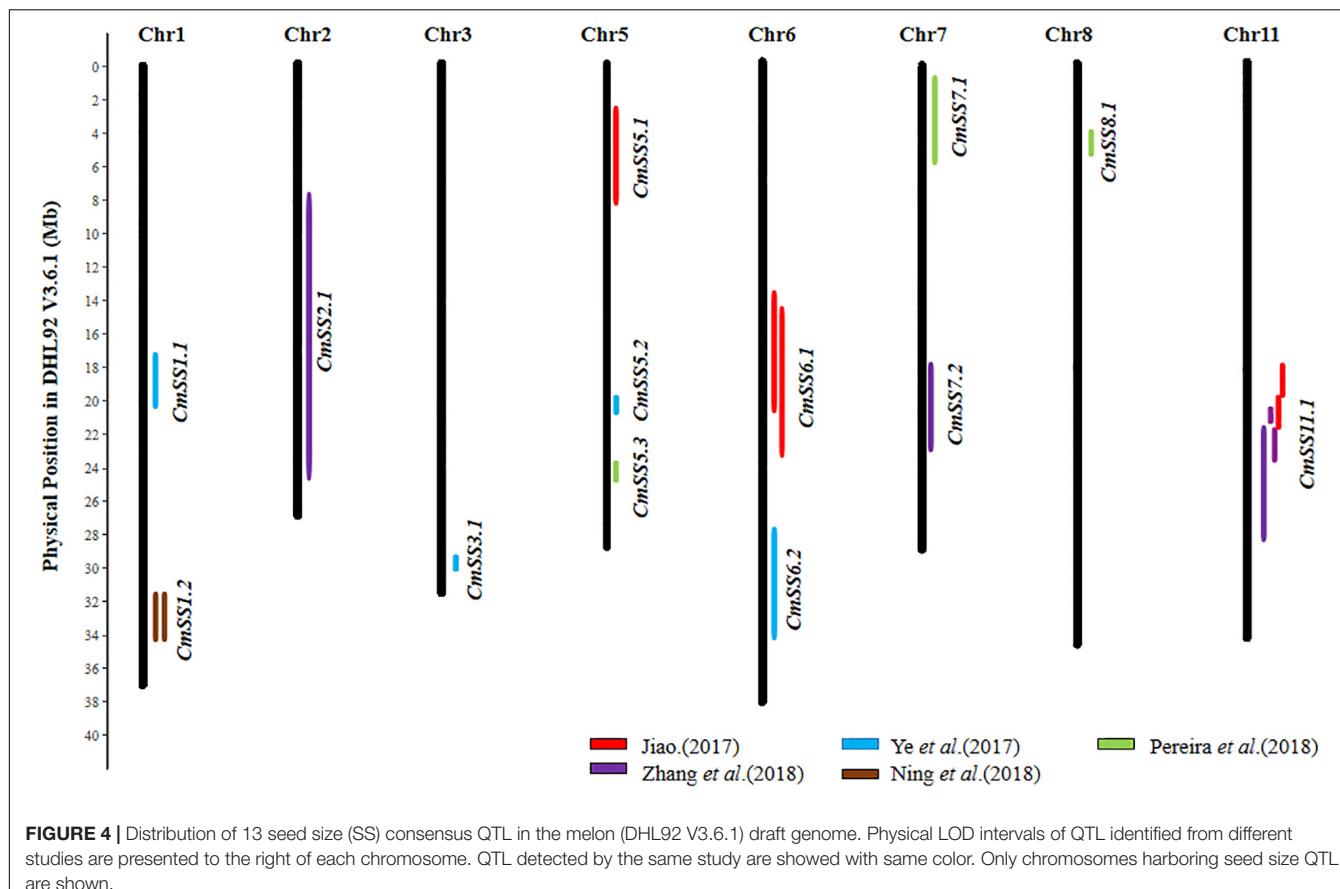
FIGURE 4 | Distribution of 13 seed size (SS) consensus QTL in the melon (DHL92 V3.6.1) draft genome. Physical LOD intervals of QTL identified from different studies are presented to the right of each chromosome. QTL detected by the same study are showed with same color. Only chromosomes harboring seed size QTL are shown.

Supplementary Table S4. Since the limited number of seed size-related QTL were detected, no consensus QTL could be inferred in pumpkin/squash.

Seed Size QTL in Cucumber

A few studies have been conducted in cucumber for QTL mapping of seed size-related traits. Chen et al. (2012) suggested that SL in cucumber is a quantitative trait controlled by multiple genes. With a RIL population developed from the cross between two cultivated cucumber lines: the large seeded 'D06157' and small seeded 'D0603,' they detected 6 SL QTL. In the RIL population derived from PI 183967 (wild) × 931 (cultivated), Wang et al. (2014) detected 14 SS QTL on five chromosomes (Chr2, 3, 4, 5, and 6) with mainly additive effect including 6 for SL (PVE = 7.5–15.6%), 4 for SWD (PVE = 7.7–18.8%), and 4 for 100SWT (PVE = 11.0–28.3%). However, only *SL5.1*, *SL6.1*, and *100SWT6.1* could be detected in two seasons. In another study, using an $F_{2:3}$ population derived from the cross between two US processing cucumber inbred lines '2A' and 'Gy8,' Lietzow (2015) identified 8 QTL for SS and 50SWT (50 seed weight) with two major QTL (PVE = 11.9% for both) located on Chr4.

From the 28 QTL identified in cucumber (details in **Supplementary Table S5**), 9 consensus SS QTL could be established, which are listed in **Supplementary Table S6**. Their chromosomal locations are graphically presented in **Figure 3**. Five of the 9 consensus SS QTL, *CsSS2.2*, *CsSS3.2*, *CsSS4.1*,

CsSS5.1, and *CsSS6.1* were identified with populations derived from crosses between the wild and cultivated cucumbers, of which *CsSS4.1* and *CsSS5.1* were detected in all studies with segregating populations created by parental materials with different seed sizes and origins. Therefore, *CsSS4.1* and *CsSS5.1* may be the potential target regions for selection during both domestication and diversifying selection.

Seed Size QTL in Melon

Five studies conducted QTL mapping for seed size in melon. Jiao (2017) conducted QTL analysis for seed size with an $F_{2:3}$ population derived from the cross of large seeded 'ms-5' and small seeded 'HM-1.' 7 QTL for SL, SWD, and 100SWT were identified on four chromosomes (Chr5, 6, 9, and 11). The 2 major-effect QTL, *Sl11.1* and *SW11.1* for SL and SWD were detected in the same region on Chr11 with PVE of 17.5 and 19.5%, respectively. With a BC_1 population from 'MR-1' (long seed) × 'M1-15' (short/narrow seed), Ye et al. (2017) detected 4 QTL, including 2 for SL, 1 for SWD and 1 for 100SWT. The 2 QTL (*SL5.1* and *SD5.1*) on Chr5 (PVE = 9.9–11.2%), and 2 QTL (*SL1.1* and *TGW1.2*) on Chr1 were validated. In addition, using a BC_1 population, Ning et al. (2018) also detected QTL for SL and 100SWT on Chr1 (LG1), but not in the same region reported by Ye et al. (2017). With a high-density genetic linkage map constructed by a RIL population, Pereira et al. (2018) identified 4 QTL for SWD with 1 major QTL (*SWQU8.1*) located on Chr8
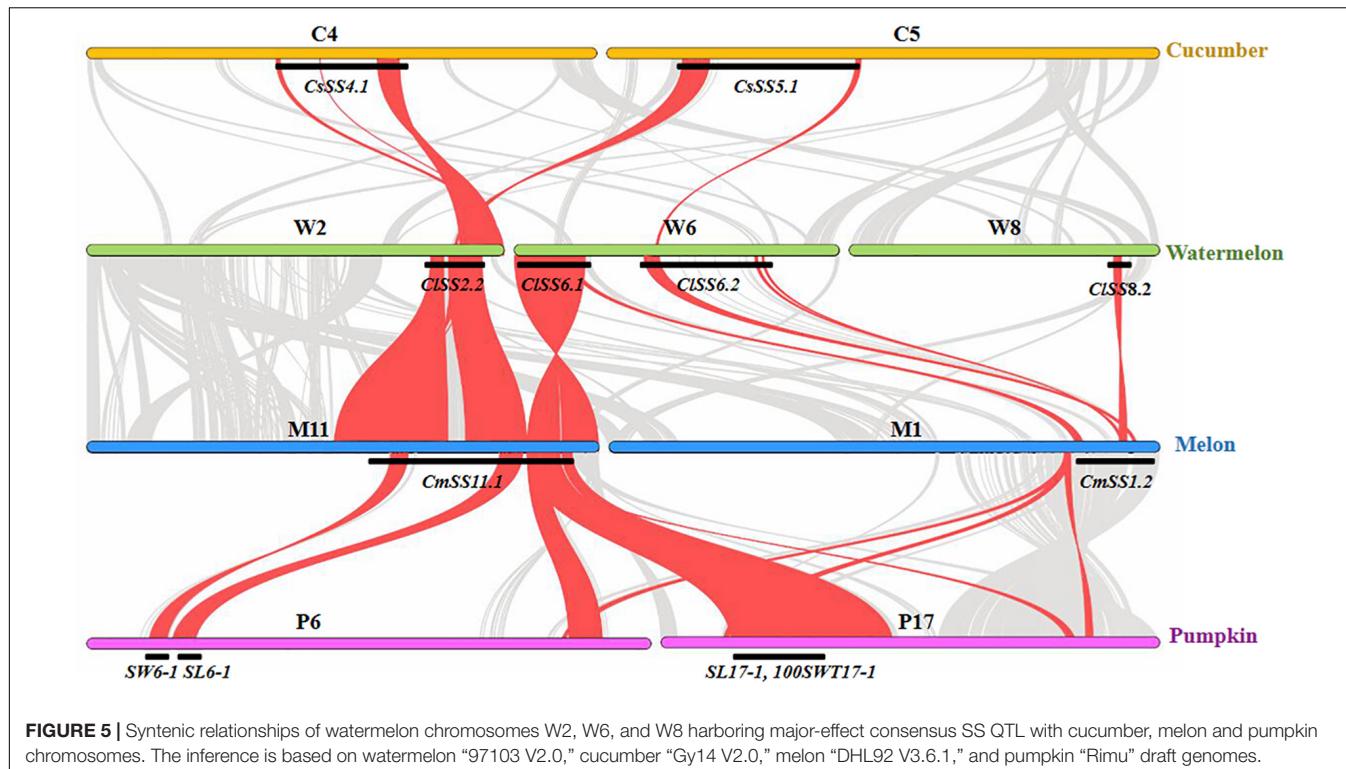
FIGURE 5 | Syntenic relationships of watermelon chromosomes W2, W6, and W8 harboring major-effect consensus SS QTL with cucumber, melon and pumpkin chromosomes. The inference is based on watermelon "97103 V2.0," cucumber "Gy14 V2.0," melon "DHL92 V3.6.1," and pumpkin "Rimu" draft genomes.

with PVE of 19.9%. Zhang et al. (2018) mapped 6 QTL on SL, SWD and 100SWT in a $F_{2:3}$ population from the cross between cantaloupe line 'ms-5' with large seed and muskmelon line 'HM1-1' with small seed, the major QTL (*sl11.1* and *sd2.1*) for SL and SWD were identified on Chr11 and Chr2 (PVE = 17.5 and 18.1%), respectively. Three QTL (*sl11.1*, *sl11.2,* and *100swt11.1*) for SL and 100SWT were detected in the same Chr11 region.

From 23 QTL detected in the six studies (**Supplementary Table S7**), 13 consensus SS QTL could be inferred, and the details are provided in **Supplementary Table S8**, and their chromosomal locations are illustrated in **Figure 4**. Both Jiao (2017) and Zhang et al. (2018) detected the QTL for SL, SWD, and 100SWT in the same region on Chr11, and other QTL were not detected reproducibly. Therefore, we speculated that the *Cm11.1* may play a major role in regulating seed size/weight variation in melon.

STRUCTURE AND FUNCTION CONSERVATION OF SS QTL IN MAJOR CUCURBITS

Previous studies have revealed high degree of sequence conservation and synteny across different cucurbit genomes, and the syntenic relationships among different chromosomal regions have also been well-established (e.g., Li et al., 2011; Garcia-Mas et al., 2012; Guo et al., 2013; Yang et al., 2014; Zhu et al., 2016; Pan et al., 2020). Comparison of consensus SS QTL locations identified from the present study may help reveal possible conservation of the genetic basis of seed size variation across major cucurbit crops. In total, 14, 9,

and 13 consensus SS QTL were established based on their physical positions in the watermelon, cucumber, and melon draft genomes, respectively (**Figures 2–4**). Among the four major cucurbits, watermelon has the most extensive studies on QTL mapping of seed size related traits (**Figure 2** and **Supplementary Tables S2, S3**). Four watermelon SS QTL have been repeatedly detected in multiple populations and environments, including *ClSS2.2*, *ClSS6.1*, *ClSS6.2*, and *ClSS8.2* (**Figure 2**). To further investigate the syntenic relationships of these major-effect, highly stable SS consensus QTL with those in other cucurbits, we aligned watermelon chromosomes 2 (W2), 6 (W6) and 8 (W8) with cucumber, melon and pumpkin genomes. We looked into the syntenic relationships of the three watermelon chromosomes (W2, W6, and W8) in cucumber, melon and pumpkin/squash. We searched the syntenic blocks based on the physical positions of these consensus QTL in the cucurbits draft genomes, which could be conveniently conducted through the cucurbit genome database at http://cucurbitgenomics.org/. The syntenic relationships of the three watermelon chromosomes (W2, W6, and W8) in cucumber, melon and pumpkin/squash were calculated by the One Step MCScanx function, and the syntenic blocks were drawn by the visualization tools Dual Systeny Pltter function in TBtools software. The syntenic relationships of the three watermelon chromosomes with those in other three cucurbits are illustrated in **Figure 5**. The syntenic blocks harboring the four watermelon SS QTL and those in melon, cucumber, and pumpkin are highlighted in red. From **Figure 5**, it is easy to see that the chromosomal region harboring watermelon *ClSS2.2* (W2) was syntenic to the regions where the cucumber *CsSS4.1* (C4) and

CsSS5.1 (C5), melon *CmSS11.1* (M11) and pumpkin *SL6-1*, *SW6-1* (P6), were respectively located. *CmSS11.1* (M11) and *CsSS4.1* (C4) were also in syntenic regions. Similarly, *ClSS6.1* (W6) was syntenic to the regions *CmSS11.1* (M11) in melon and pumpkin *SL17-1*, *100SWT17-1* (P17). *CmSS1.2* (M1) and the regions harboring both *ClSS6.1* (W6) and *ClSS8.2* (W8) were syntenic. These observations suggested possible conservation of functions and structure genes/QTL for seed size control among cucurbit crops. Such conservation could also be possible for QTL in other regions. However, more QTL mapping work is needed to make such inferences.

The molecular mechanisms of seed size regulation in cucurbits are unknown. The functions of many genes for organ development are high conserved (for example, fruit size and shape as reviewed in Pan et al., 2020). From QTL mapping studies, the four watermelon consensus SS QTL (*ClSS2.2*, *ClSS6.1*, *ClSS6.2*, and *ClSS8.2*) have been delimited to relatively small regions. We examined the annotated genes in the four regions from the watermelon draft genomes. In the watermelon 97103 V2.0 draft genome, 21, 11, 6, and 34 genes were predicted in the *ClSS2.2*, *ClSS6.1*, *ClSS6.2* and *ClSS8.2* consensus QTL regions, respectively, which are listed in **Supplementary Table S9**. Among those genes, several seem to be good candidates for these SS QTL. As discussed early (see section "Introduction," **Supplementary Table S1**), the ubiquitination pathway may play important roles in control of seed size (e.g., Song et al., 2007; Li et al., 2010; Su et al., 2011; Xia et al., 2013; Wang et al., 2016; Dong et al., 2017). From our recent study in pumpkin, the E3 ubiquitin-protein ligase gene was located in the *SL6-1* and *SW6-1* major-effect QTL region (Wang et al., 2019). In rice, a gene encoding a zinc finger protein has been verified in control grain weight (Huang et al., 2011). In the watermelon SS QTL regions, there were six genes with predicted functions to encode zinc finger proteins including *Cla97C02G045430*, *Cla97C02G045460* and *Cla97C02G045500* in the *ClSS2.2* region, *Cla97C06G114460* in *ClSS6.1* region, and *Cla97C08G154340* and *Cla97C08G154360* in *ClSS8.2* region. Another gene, *Cla97C08G154570* encoding E3 ubiquitin-protein ligase was located in *ClSS8.2* region (**Supplementary Table S9**). These genes might be interesting candidates for the seed size QTL that merit consideration in future studies.

CONCLUSION REMARKS

In this study, we summarized QTL for seed size-related traits in major cucurbits, and identified consensus SS QTL in watermelon, melon and cucumber. Many of these SS QTL were located in non-syntenic regions in the three cucurbit crops. In watermelon and melon, it seems that several stable consensus QTL are located in syntenic blocks in different cucurbit crops (**Figure 5**), which may suggest shared common QTL for seed size control in some genetic backgrounds. It should be noted that QTL mapping studies for seed size/weight are very limited in major cucurbit crops (none in minor cucurbits). Many of these consensus QTL were inferred from single study. The genomic regions harboring these QTL are still very large. The effects of these QTL need to be confirmed, and their locations need to be refined in future studies. Seed size/weight was the target of selection during domestication from wild ancestors, which usually have small seeds. In seed-use cucurbits like watermelon and pumpkin, seed size may be under further selection during breeding, while in melon and cucumber where flesh (endocarp) is consumed, seed size may not be a main target in breeding. Understanding the genetic basis and the roles of selection in diversifying and domestication shaping seed size variation might be some interesting topics in future studies. Such information is also important in marker-assisted selection in breeding for seed size in cucurbit crops.

AUTHOR CONTRIBUTIONS

YGu and MG conducted literature review and wrote the manuscript. XXL, MX, and XSL performed the comparative analysis. FL and SQ helped review of watermelon, melon, and pumpkin data. YZ, JL, XJL, and YGa analyzed and reviewed cucumber and melon data. All authors read and approved the final submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00304/full#supplementary-material

REFERENCES

Al-Khalifa, A. S. (1996). Physicochemical characteristics, fatty acid composition, and lipoxygenase activity of crude pumpkin and melon seed oils. *J. Agric. Food Chem.* 44, 964–966. doi: 10.1021/jf950519s

Assefa, T., Otyama, P. I., Brown, A. V., Kalberer, S. R., Kulkarni, R. S., and Cannon, S. B. (2019). Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics* 20:527. doi: 10.1186/s12864-019-5907-7

Baboli, Z. M., and Kordi, A. A. S. (2010). Characteristics and composition of watermelon seed oil and solvent extraction parameters effects. *J. Am. Oil Chem. Soc.* 87, 667–671. doi: 10.1007/s11746-010-1546-5

Bai, X., Luo, L., Yan, W., Kovi, M. R., and Xing, Y. (2010). Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus *qGL7*. *BMC Genet.* 11:16. doi: 10.1186/1471-2156-11-16

Charlotte, M., Rachel, W., Neil, M., Martin, T., Joshua, B., Abdelhak, F., et al. (2019). Variation in expression of the *HECT E3 ligase UPL3* modulates *LEC2*

levels, seed size, and crop yields in *Brassica napus*. *Plant Cell* 31, 2370–2385. doi: 10.1101/334581

Chaudhury, A. M., Koltunow, A., Payne, T., Luo, M., Tucker, M. R., Dennis, E. S., et al. (2001). Control of early seed development. *Annu. Rev. Cell Dev. Biol.* 17, 677–699. doi: 10.1146/annurev.cellbio.17.1.677

Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., et al. (2016). Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* 171, 606–622. doi: 10.1104/pp.15.01992

Chen, L. L., Qin, Z. W., Zhou, X. Y., Xin, M., and Wu, T. (2012). Molecular marker and genetic analysis of cucumber seed length. *Chinese Sci. Bull.* 28, 165–170.

Chen, N. L., Gao, Y. M., and Zhang, Y. X. (2006). Effects of seed size and light intensity on seedling growth of watermelon. *China Cucurbits Vegetables* 1–3. doi: 10.3969/j.issn.1673-2871.2006.03.001

Chettoor, A. M., Phillips, A. R., Coker, C. T., Dilkes, B., and Evans, M. M. S. (2016). Maternal gametophyte effects on seed development in maize. *Genetics* 204, 233–248. doi: 10.1534/genetics.116.191833

Coomes, D. A., and Grubb, P. J. (2003). Colonization, tolerance, competition and seed-size variation within functional groups. *Trends Ecol. Evol.* 18, 283–291. doi: 10.1016/S0169-5347(03)00072-7

Dong, H., Dumenil, J., Lu, F. H., Na, L., Vanhaeren, H., Naumann, C., et al. (2017). Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in *Arabidopsis*. *Genes Dev.* 31, 197–208. doi: 10.1101/gad.292235.116

Downes, B. P., Stupar, R. M., Gingerich, D. J., and Vierstra, R. D. (2003). The HECT ubiquitin-protein ligase (UPL) family in Arabidopsis: UPL3 has a specific role in trichome development. *Plant J.* 35, 729–742. doi: 10.1046/j.1365-313X.2003.01844.x

Du, J., Wang, S., He, C., Zhou, B., Ruan, Y. L., and Shou, H. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68, 1955–1972. doi: 10.1093/jxb/erw460

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Figueiredo, D., and Kohler, C. (2014). Signaling events regulating seed coat development. *Biochem. Soc. Trans.* 42, 358–363. doi: 10.1042/BST20130221

Gao, M. L., Zhao, F. F., Wang, Y. S., Yuan, C. Z., Wei, X. M., Li, J. Y., et al. (2016). Genetic analysis and correlation of seed traits in mini watermelon. *Jiangsu Agric. Sci.* 44, 259–261.

Garcia, D., Gerald, J. N. F., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* 17, 52–60. doi: 10.1105/tpc.104.027136

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G., and Berger, F. (2003). *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol.* 131, 1661–1670. doi: 10.1104/pp.102.018762

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V. M., et al. (2012). The genome of melon (*Cucumis melo* L.). *Proc. Natl Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109

Gómez, J. M. (2004). Bigger is not always better: conflicting selective pressure on seed size in *Quercus ilex*. *Evolution* 58, 71–80. doi: 10.1111/j.0014-3820.2004.tb01574.x

Gu, Y. Z., Li, W., Jiang, H. W., Wang, Y., Gao, H. H., Liu, M., et al. (2017). Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* 68, 2717–2729. doi: 10.1093/jxb/erx147

Guo, S. G., Zhang, J. G., Sun, H. H., Salse, J., Lucas, W., Zhang, H. Y., 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

Hu, Z. J., Lu, S. J., Wang, M. J., He, H. H., Sun, L., Wang, H. R., et al. (2018). A novel QTL *qTGW3* encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. *Mol. Plant* 11, 736–749. doi: 10.1016/j.molp.2018.03.005

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

Huang, X. H., Zhao, Y., Wei, X. H., Li, C. Y., Wang, A. H., Zhao, Q., et al. (2011). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* 44, 32–39. doi: 10.1038/ng.1018

Huang, Y., Bai, X. F., Cheng, N. N., Xiao, J. H., Li, X. H., and Xing, Y. Z. (2019). Wide Grain 7 increases grain width by enhancing H3K4me3 enrichment in the *OsMADS1* promoter in rice (*Oryza sativa* L.). *Plant J.* [Epub ahead of print].

Jiang, W. B., Huang, H. Y., Hu, Y. W., Zhu, S. W., Wang, Z. Y., and Lin, W. H. (2013). Brassinosteroid regulates seed size and shape in *Arabidopsis*. *Plant Physiol.* 162, 1965–1977. doi: 10.1104/pp.113.217703

Jiao, S. Q. (2017). *QTL Analysis of Fruit and Seed Related Traits in Melon*. Harbin: Northeast Agricultural University, 4–25.

Jofuku, K. D., Omidyar, P. K., Gee, Z., and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3117–3122. doi: 10.1073/pnas.0409893102

Johnson, C. S., Kolevski, B., and Smyth, D. R. (2002). *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell* 14, 1359–1375. doi: 10.1105/tpc.001404

Kanako, K., Shigeru, K., Katsuyuki, O., Yuki, A., Tsuyu, A., Izumi, K., et al. (2010). A novel kinesin 13 protein regulating rice seed length. *Plant Cell Physiol.* 51, 1315–1329. doi: 10.1093/pcp/pcq092

Kim, K. H., Hwang, J. H., Han, D. Y., Park, M., Kim, S., Choi, D., et al. (2015). Major quantitative trait loci and putative candidate genes for powdery mildew resistance and fruit-related traits revealed by an intraspecific genetic map for watermelon (*Citrullus lanatus* var. *lanatus*). *PLoS One* 10:e0145665. doi: 10.1371/journal.pone.0145665

Li, D. W., Cuevas, H. E., Yang, L., Li,Y. H., Garcia-Mas, J., Zalapa, J., et al. (2011). Syntenic relationships between cucumber (*Cucumis sativus* L.) and melon (*C. melo* L.) chromosomes as revealed by comparative genetic mapping. *BMC Genomics* 12:396. doi: 10.1186/1471-2164-12-396

Li, M. R., Li, X. X., Zhou, Z. J., Wu, P. Z., Fang, M. C., Pan, X. P., et al. (2016). Reassessment of the four yield-related genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* 7:377. doi: 10.3389/fpls.2016.00377

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Li, N., Shang, J. L., Wang, J. M., Zhou, D., Li, N. N., and Ma, S. W. (2018a). Fine mapping and discovery of candidate genes for seed size in watermelon by genome survey sequencing. *Sci. Rep.* 8, 843–854. doi: 10.1038/s41598-018-36104-w

Li, N., Xu, R., Duan, P. G., and Li, Y. H. (2018b). Control of grain size in rice. *Plant Rep.* 31, 237–251. doi: 10.1007/s00497-018-0333-6

Li, N., Xu, R., and Li, Y. H. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70, 435–463. doi: 10.1146/annurev-arplant-050718-095851

Li, Q., Li, H. B., Huang, W., Xu, Y. C., Zhou, Q., Wang, S. H., et al. (2019). A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). *Gigascience* 8:giz072. doi: 10.1093/gigascience/giz072

Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight. *BMC Plant Biol.* 10:143. doi: 10.1186/1471-2229-10-143

Li, Y. D., Zhang, S. L., Dong, R. Z., Wang, L., Yao, J., Van, N. S., et al. (2019). The grapevine homeobox gene *VvHB58* influences seed and fruit development through multiple hormonal signaling pathways. *BMC Plant Biol.* 19:523. doi: 10.1186/s12870-019-2144-9

Li, Y. H., Zheng, L. Y., Corke, F., Smith, C., and Bevan, M. (2008). Control of final seed and organ size by the *DA1* gene family in *Arabidopsis thaliana*. *Genes Dev.* 22, 1331–1336. doi: 10.1101/gad.463608

Lietzow, C. D. (2015). *Characterization and QTL Mapping of Parthenocarpic Fruit Set in Processing Cucumber (Cucumis sativus* L.). (Ann Arbor, MI: ProQuest LLC), 157–158.

Liu, C. Q., Gao, P., and Luan, F. S. (2014). Construction of a genetic linkage map and QTL analysis of fruit-associated traits in watermelon. *Sci. Agric. Sin.* 47, 2814–2829.

Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., and Chaudhury, A. (2005). MINISEED3 (*MINI3*), a WRKY family gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (*LRR*) *KINASE* gene, are regulators of seed size in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17531–17536. doi: 10.1073/pnas.0508418102

Meru, G., Fu, Y., Leyva, D., Sarnoski, P., and Yagiz, Y. (2018). Phenotypic relationships among oil, protein, fatty acid composition and seed size traits in *Cucurbita pepo*. *Sci. Hortic.* 233, 47–53. doi: 10.1016/j.scienta.2018.01.030

Meru, G., and McGregor, C. (2013). Genetic mapping of seed traits correlated with seed oil percentage in watermelon. *Hortscience* 48, 955–959. doi: 10.21273/HORTSCI.48.8.955

Mizukami, Y., and Fischer, R. L. (2000). Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 942–947. doi: 10.1073/pnas.97.2.942

Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B., and Westoby, M. (2005). A brief history of seed size. *Science* 307, 576–580. doi: 10.1126/science.1104863

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., et al. (2018). *De novo* assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnol. J.* 16, 1161–1171. doi: 10.1111/pbi.12860

Ning, X. F., Wang, X. L., Yu, Z. J., Lu, S. M., and Li, G. (2018). Inheritance and molecular mapping of tight-placenta gene and seed traits in Chinese Hami Melon 'Queen'. *Hortscience* 53, 788–794. doi: 10.21273/HORTSCI13046-18

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., et al. (2005). Functional genomic analysis of the *AUXIN RESPONSE FACTOR* gene family members in *Arabidopsis thaliana*: unique and overlapping functions of *ARF7* and *ARF19*. *Plant Cell* 17, 444–463. doi: 10.1105/tpc.104.028316

Pan, Y. P., Wang, Y. H., McGregor, C., Liu, S., Luan, F. S., Gao, M. L., 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

Paris, H. S., and Nerson, H. (2003). Seed dimensions in the subspecies and cultivar-groups of *Cucurbita pepo*. *Gen. Resour. Crop Evol.* 50, 615–625. doi: 10.1023/a: 1024464831595

Pereira, L., Ruggieri, V., Pérez, S., Alexiou, K. G., Fernández, M., Jahrmann, T., et al. (2018). QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biol.* 18:324. doi: 10.1186/s12870-018-1537-5

Poole, C. F., Grimball, P. C., and Porter, D. R. (1941). Inheritance of seed characters in watermelon. *J. Agric. Res.* 63, 433–456.

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: 10.4161/psb.5.7.12095

Prothro, J., Sandlin, K., Abdel-Haleem, H., Bachlava, E., and Mcgregor, C. (2012a). Main and epistatic quantitative trait loci associated with seed size in watermelon. *J. Am. Soc. Hortic. Sci.* 137, 452–457. doi: 10.21273/JASHS.137.6.452

Prothro, J., Sandlin, K., Gill, R., Bachlava, E., White, V., and Knapp, S. (2012b). Mapping of the *Egusi* seed trait locus (*eg*) and quantitative trait Loci associated with seed oil percentage in watermelon. *J. Am. Soc. Hortic. Sci.* 137, 311–315. doi: 10.21273/JASHS.137.5.311

Ren, D. Q., Wang, X. C., Yang, M., Yang, L., He, G. M., and Deng, X. W. (2019). A new regulator of seed size control in *Arabidopsis* identified by a genome-wide association study. *New Phytol.* 222, 895–906. doi: 10.1111/nph.15642

Ren, Y., Mcgregor, C., Zhang, Y., Gong, G., Zhang, H., Guo, S. G., 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

Ren, Y., Zhao, H., Kou, Q. H., Jiang, J., Guo, S. G., Zhang, H. Y., et al. (2012). A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PLoS One* 7:e29453. doi: 10.1371/journal.pone.0029453

Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., et al. (2018). An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. *Sci. Rep.* 8:8088. doi: 10.1038/s41598-018-26416-2

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

Savadi, S. (2018). Molecular regulation of seed development and strategies for engineering seed size in crop plants. *Plant Growth Regul.* 84, 401–422. doi: 10.1007/s10725-017-0355-3

Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2005). The *AUXIN RESPONSE FACTOR* 2 gene of *Arabidopsis* links auxin

signaling, cell division, and the size of seeds and other organs. *Development* 133, 251–261. doi: 10.1242/dev.02194

Segami, S., Kono, I., Ando, T., Yano, M., Kitano, H., Miura, K., et al. (2012). Small and round seed 5, gene encodes alpha-tubulin regulating seed cell elongation in rice. *Rice* 5:4. doi: 10.1186/1939-8433-5-4

Segami, S., Takehara, K., Yamamoto, T., Kido, S., and Miura, K. (2017). Overexpression of *SRS5* improves grain size of brassinosteroid-related dwarf mutants in rice (*Oryza sativa* L.). *Breed. Sci.* 67, 393–397. doi: 10.1270/jsbbs.16198

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

Si, L. Z., Chen, J. Y., Huang, X. H., Gong, H., Luo, J. H., Hou, Q. Q., et al. (2016). *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Silvertown, J. W. (1981). Seed size, life span, and germination date as coadapted features of plant life history. *Am. Nat.* 118, 860–864. doi: 10.2307/2460817

Smitchger, J., and Weeden, N. F. (2018). The ideotype for seed size: a model examining the relationship between seed size and actual yield in pea. *Int. J. Agron.* 2018:9658707. doi: 10.1155/2018/9658707

Song, X. J., Huang, W., Shi, M., Zhu, M. Z., and Lin, H. X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type *E3* ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Stanton, M. L. (1984). Seed variation in wild radish: effect of seed size on components of seedling and adult fitness. *Ecology* 65, 1105–1112. doi: 10.2307/1938318

Su, Z. Q., Hao, C. Y., Wang, L. F., Dong, Y. C., and Zhang, X. Y. (2011). Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Sun, H. H., Wu, S., Zhang, G. Y., Jiao, C., Guo, S. G., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid, cucurbita, genomes. *Mol. Plant* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003

Sun, X. D., Shantharaj, D., Kang, X. J., and Ni, M. (2010). Transcriptional and hormonal signaling control of Arabidopsis seed development. *Curr. Opin. Plant Biol.* 13, 611–620. doi: 10.1016/j.pbi.2010.08.009

Tan, X. Z., Ge, Y., Xu, W. L., Cui, C. S., and Qu, S. P. (2013). Construction of genetic linkage map and QTL analysis for seed width in pumpkin (*Cucurbita Maxima*). *Acta Bot. Borealioccidentalia Sin.* 33, 697–702.

Tanaka, T., Wimol, S., and Mizutani, T. (1995). Inheritance of fruit shape and seed size of watermelon. *J. Jpn. Soc. Hort. Sci.* 64, 543–548. doi: 10.2503/jjshs.64.543

Tian, Y. F., Wei, Y. Y., Chang, X. K., Yang, M. K., Wu, W. J., and Wei, X. (2012). Effects of seed size on germinating period and seedling quality of cucumber, watermelon and squash. *Shandong Agric. Sci.* 44, 40–42. doi: 10.14083/j.issn.1001-4942.2012.08.039

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., et al. (2017). Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* 24, 51–58. doi: 10.1093/dnares/dsw047

Wang, A. H., Garcia, D., Zhang, H. Y., Feng, K., Chaudhury, A., Berger, F., et al. (2010). The VQ motif protein *IKU1* regulates endosperm growth and seed size in *Arabidopsis*. *Plant J.* 63, 670–679. doi: 10.1111/j.1365-313X.2010.04271.x

Wang, M., Miao, H., Zhang, S. P., Liu, S. L., Dong, S. Y., Wang, Y., et al. (2014). Inheritance analysis and QTL mapping of cucumber seed size. *Acta Hortic. Sin.* 41, 63–72. doi: 10.16420/j.issn.0513-353x.2014.01.010

Wang, P., Zhang, Y., Bai, L. H., Liu, J. C., and Yang, W. X. (2012). Correlation and path analysis about yield component factors of seed-used pumpkin (*Cucurbita pepo* L.). *Northern Hortic.* 36, 16–19.

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* 44, 950–954. doi: 10.1038/ng.2327

Wang, S. K., Li, S., Liu, Q., Wu, K., Zhang, J. Q., Wang, S. S., et al. (2015). The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47, 949–954. doi: 10.1038/ng.3352

Wang, Y. H., Bo, K. L., Gu, X. F., Pan, J. S., Li, Y. H., Chen, J. F., 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, Y. H., Wang, C. J., Han, H. Y., Luo, Y. S., Wang, Z. C., Yan, Y. L., et al. (2019). Construction of a high-density genetic map and analysis of seed-related traits using specific length amplified fragment sequencing for *Cucurbita maxima*. *Front. Plant Sci.* 10:1782. doi: 10.3389/fpls.2019.01782

Wang, Z., Li, N., Jiang, S., Gonzalez, N., Huang, X., Wang, Y., et al. (2016). SCFSAP controls organ size by targeting PPD proteins for degradation in *Arabidopsis thaliana*. *Nat. Commun.* 7:11192. doi: 10.1038/ncomms11192

Wu, G. L., Du, G. Z., and Shang, Z. H. (2006). Contribution of seed size and its fate to vegetation renewal: a review. *Chin. J. Appl. Ecol.* 17, 1969–1972. doi: 10.13287/j.1001-9332.2006.0387

Wu, S., Shamimuzzaman, M., Sun, H., Salse, J., Sui, X., Wilder, A., et al. (2017). The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. *Plant J.* 92, 963–975. doi: 10.1111/tpj.13722

Xia, D., Zhou, H., Liu, R., Dan, W., Li, P., Wu, B., et al. (2018). GL3.3, a novel QTL encoding a *GSK3/SHAGGY*-like kinase, epistatically interacts with GS3, to form extra-long grains in rice. *Mol. Plant* 11, 754–756. doi: 10.1016/j.molp.2018.03.006

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

Xiang, C. P., Shi, X. M., and Zhang, Y. (2002). Effect of seed size on seed quality and yield of bitter gourd. *J. Changjiang Vegetables* 35–36.

Xiao, Y. G., Sun, Q. B., Kang, X. J., Chen, C. B., and Ni, M. (2016). *SHORT HYPOCOTYL UNDER BLUE1* or *HAIKU2* mixepression alters canola and Arabidopsis seed development. *New Phytol.* 209, 636–649. doi: 10.1111/nph.13632

Xie, D. S., Xu, Y. C., Wang, J. P., Liu, W. R., Zhou, Q., Luo, S., et al. (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, D. H., Zhang, Y. W., Li, C. G., and Cui, C. S. (2006). Combining ability of economic characters in seed pumpkin (*Cucurbita maxima* Duch.). *China Vegetables* 01, 13–15.

Yang, L. M., Koo, D. H., Li, D. W., Zhang, T., Jiang, J. M., Luan, F. S., et al. (2014). Next-generation sequencing. FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in Cucumis. *Plant J.* 77, 16–30. doi: 10.1111/tpj.12355

Yang, L. M., Koo, D. H., Li, Y. H., Zhang, X. J., Luan, F. S., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313x.2012.05017.x

Yang, X., Wu, F., Lin, X., Du, X., Chong, K., Gramzow, L., et al. (2012). Live and let die-the Bsister MADS-box gene *OsMADS29* controls the degeneration of cells in maternal tissues during seed development of rice (*Oryza sativa*). *PLoS One* 7:e51435. doi: 10.1371/journal.pone.0051435

Ye, W. Z., Liu, S., Ma, H. Y., Liu, H. Y., Li, G. Y., and Luan, F. S. (2017). QTL analysis on related seed traits in melon based on CAPS markers. *North. Hortic.* 12, 119–128. doi: 10.11937/bfyy.201712028

Yongjae, K., Taejin, Y., Younghoon, P., Yongjik, L., Suncheol, K., Yongkwon, K., et al. (2009). Development of near isogenic lines with various seed sizes and study on seed size-related characteristics in watermelon. *Korean J. Breed. Sci.* 41, 403–411.

Zeng, Y. F., Wen, J. Y., Zhao, W. B., Wang, Q., and Huang, W. C. (2020). Rational improvement of rice yield and cold tolerance by editing the three genes *OsPIN5b*, *GS3*, and *OsMYB30* with the CRISPR-Cas9 system. *Front. Plant Sci.* 10:1633. doi: 10.3389/fpls.2019.01663

Zhang, G. F., and Zhang, J. N. (2011). Inheritance of seed size in watermelon. *Jiangsu Agric. Sci.* 39, 216–217. doi: 10.15889/j.issn.1002-1302.2011.04.170

Zhang, J. (1996). Inheritance of seed size from diverse crosses in watermelon. *Cucurbit Genet. Coop. Rep.* 19, 67–69.

Zhang, K. X., Dai, D. Y., Wang, H. N., Yu, M. Y., and Sheng, Y. Y. (2018). Genetic and QTL analysis of seed traits in melon (*Cucumis melon* L.). *Acta Agric. Zhejiang.* 30, 1496–1503. doi: 10.1186/1471-2229-11-111

Zhang, X., Rhodes, B., and Wang, M. (1994). "Genes controlling watermelon seed size," in *Cucurbitaceae '94: Evaluation and Enhancement of Cucurbit Germplasm*, eds G. Lester, and J. Dunlap, (Alexandria, VA: ASHS Press), 144–147.

Zhang, X. J., Jiang, W. B., and Pang, Y. Z. (2016). Advances in the regulation mechanism of plant seed size. *Plant. Physiol. J.* 998–1010. doi: 10.13592/j.cnki.ppj.2016.0159

Zhou, H. W., Lu, B. Y., Ma, H. Y., Gao, P., Luan, F. S., Gao, Q. F., et al. (2016). QTL mapping of watermelon seed traits. *Acta Hortic. Sin.* 43, 715–723. doi: 10.16420/j.issn.0513-353x.2015-0973

Zhou, Y., Zhang, X. J., Kang, X. J., Zhao, X. Y., Zhang, X. S., and Ni, M. (2009). *SHORT HYPOCOTYL UNDER BLUE1* associates with *MINISEED3* and *HAIKU2* promoters in vivo to regulate *Arabidopsis* seed development. *Plant Cell* 21, 106–117. doi: 10.2307/40537464

Zhu, H. Y., Song, P. Y., Koo, D. H., Guo, L. Q., Li, Y. M., Sun, S. R., et al. (2016). Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genomics* 17:557. doi: 10.1186/s12864-016-2870-4

Zhu, Z. G., Liang, H. L., Chen, G. P., Li, F. F., Wang, Y. Y., Liao, C. G., et al. (2019). The bHLH transcription factor SlPRE2 regulates tomato fruit development and modulates plant response to gibberellin. *Plant Cell Rep.* 38, 1053–1064. doi: 10.1007/s00299-019-02425-x

A Major QTL Located in Chromosome 8 of *Cucurbita moschata* Is Responsible for Resistance to *Tomato Leaf Curl New Delhi Virus*

Cristina Sáez[1], Cecilia Martínez[2], Javier Montero-Pau[3], Cristina Esteras[1], Alicia Sifres, José Blanca[1], María Ferriol[4], Carmelo López[1] and Belén Picó[1]*

[1] Institute for the Conservation and Breeding of Agricultural Biodiversity, Universitat Politècnica de València, Valencia, Spain, [2] Agrifood Campus of International Excellence (ceiA3), Department of Biology and Geology, Universidad de Almería, Almería, Spain, [3] Department of Biochemistry and Molecular Biology, Universitat de València, Valencia, Spain, [4] Instituto Agroforestal Mediterráneo, Universitat Politècnica de València, Valencia, Spain

Tomato leaf curl New Delhi virus (ToLCNDV) is a bipartite whitefly transmitted begomovirus, responsible since 2013 of severe damages in cucurbit crops in Southeastern Spain. Zucchini (*Cucurbita pepo*) is the most affected species, but melon (*Cucumis melo*) and cucumber (*Cucumis sativus*) are also highly damaged by the infection. The virus has spread across Mediterranean basin and European countries, and integrated control measures are not being enough to reduce economic losses. The identification of resistance genes is required to develop resistant cultivars. In this assay, we studied the inheritance of the resistance to ToLCNDV previously identified in two *Cucurbita moschata* accessions. We generated segregating populations crossing both resistant pumpkins, an American improved cultivar Large Cheese (PI 604506) and an Indian landrace (PI 381814), with a susceptible *C. moschata* genotype (PI 419083). The analysis of symptoms and viral titers of all populations established the same monogenic recessive genetic control in both resistant accessions, and the allelism tests suggest the occurrence of alleles of the same *locus*. By genotyping with a single nucleotide polymorphism (SNP) collection evenly distributed along the *C. moschata* genome, a major quantitative trait locus (QTL) was identified in chromosome 8 controlling resistance to ToLCNDV. This major QTL was also confirmed in the interspecific *C. moschata* × *C. pepo* segregating populations, although *C. pepo* genetic background affected the resistance level. Molecular markers here identified, linked to the ToLCNDV resistance *locus*, are highly valuable for zucchini breeding programs, allowing the selection of improved commercial materials. The duplication of the candidate region within the *C. moschata* genome was studied, and genes with paralogs or single-copy genes were identified. Its synteny with the region of chromosome 17 of the susceptible *C. pepo* revealed an INDEL including interesting candidate genes. The chromosome 8 candidate region of *C. moschata* was also syntenic to the region in chromosome 11

of melon, previously described as responsible of ToLCNDV resistance. Common genes in the candidate regions of both cucurbits, with high- or moderate-impact polymorphic SNPs between resistant and susceptible *C. moschata* accessions, are interesting to study the mechanisms involved in this recessive resistance.

INTRODUCTION

Tomato leaf curl New Delhi virus (ToLCNDV) is an economically important begomovirus (family *Geminiviridae*) with two circular single-stranded DNA genome components of ∼2.7 kb, designated as DNA-A and DNA-B (Padidam et al., 1995; Jyothsna et al., 2013). ToLCNDV is transmitted in nature by the whitefly *Bemisia tabaci* biotypes MEAM1 and MED (Chang et al., 2010; Rosen et al., 2015; Janssen et al., 2017), but some isolates of this virus can also be transmitted by mechanical inoculation (Usharani et al., 2004; Chang et al., 2010; Sohrab et al., 2013; López et al., 2015).

ToLCNDV has a wide host range. It affects crops of the *Solanaceae* family, such as tomato (*Solanum lycopersicum* L.), potato (*Solanum tuberosum* L.), chili pepper (*Capsicum annuum* L.), and eggplant (*Solanum melongena* L.) (Padidam et al., 1995; Hussain et al., 2004; Usharani et al., 2004; Pratap et al., 2011). It is also highly damaging to crops of the *Cucurbitaceae* family, including luffa (*Luffa cylindrica* M. Roem.), ash gourd [*Benincasa hispida* (Thunb.) Cogn.], cucumber (*Cucumis sativus* L.), watermelon (*Citrullus lanatus* L), melon (*C. melo* L.), and different types of squashes (*Cucurbita* spp.) (Sohrab et al., 2003; Ito et al., 2008; Singh et al., 2009; Chang et al., 2010; Roy et al., 2013). Recently, it has been reported to be affecting species of other plant families, such as opium poppy (*Papaver somniferum* L., Papaveraceae) (Srivastava et al., 2016), cotton (*Gossypium hirsutum* L., Malvaceae) (Zaidi et al., 2016), soybean (*Glycine max*, Fabaceae L. Merr.) (Jamil et al., 2017), and firecracker flower (*Crossandra infundibuliformis* L. Nees, *Acanthaceae*) (Sundararaj et al., 2019). Furthermore, some weeds as black nightshade (*Solanum nigrum* L.), thorn apple (*Datura stramonium* L.), squirting cucumber [*Ecballium elaterium* (L.) A. Rich], smooth sowthistle (*Sonchus oleraceus* L.), false daisy [*Eclipta prostrata* (L.) L.], and apple of Sodom [*Calotropis procera* (Aiton) Dryand.] (Haider et al., 2006; Moriones et al., 2017; Zaidi et al., 2017; Juárez et al., 2019) have been found to be hosts of the virus, acting as reservoirs during the whole year.

ToLCNDV was first detected in North India in 1995 (Srivastava et al., 1995), from where it spread to South and Southeast Asian countries. It was limited to Asia until 2012, when it was reported affecting cucurbits in different Mediterranean countries, first in Spain (Juárez et al., 2014) and later in Tunisia (Mnari-Hattab et al., 2015), Italy (Panno et al., 2016), Morocco (Sifres et al., 2018), Greece (Orfanidou et al., 2019), and Algeria (Kheireddine et al., 2019). More recently, the virus has been identified in cucurbits plants in Portugal and Estonia (EPPO, 2019), which is indicative of the rapid spread of ToLCNDV through Europe. The most affected crop in European countries is Zucchini squash (*Cucurbita pepo* L. subsp. *pepo*). In this crop,

the virus causes severe stunting of plants, which exhibit upward and downward curling of the leaves, severe mosaic, and fruit skin roughness (Juárez et al., 2014). Infected plants often present partial or complete yield loss and fruits with lower market value. Zucchini is one of the most widely grown crops and appreciated vegetable in the Mediterranean basin. This region produced nearly 300,000 tm of this vegetable and other species of the *Cucurbita* genus (pumpkins, squash, and gourds) in FAO (2017), representing almost 24% of world production, excluding China and India. Before the arrival of ToLCNDV, the aphid-borne potyvirus *Zucchini yellow mosaic virus* (ZYMV) was the major viral pathogen of this crop (Capuozzo et al., 2017). Since 2013, ToLCNDV is the most prevalent virus in the area, where it is an important constraint to zucchini production. In the background of the severe epidemic outbreaks of ToLCNDV in cucurbits, both in greenhouses and in open fields, European and Mediterranean Plant Protection Organization (EPPO) has added this virus to the EPPO Alert List (EPPO, 2017).

Cultural practices, such as the control of the whitefly vector, the elimination of infected plants, and the avoidance of the most susceptible cultivars, are not very effective in preventing ToLCNDV outbreaks (EPPO, 2017). In fact, breeding resistant varieties is considered the most economical and effective method to control virus diseases. Genetic resistance to ToLCNDV has been identified in some accessions of the *Cucurbita* genus (Sáez et al., 2016). In that work, authors screened for ToLCNDV resistance in a large collection of *Cucurbita* spp. accessions including landraces and commercial varieties of the cultivated species (*C. pepo* L., *C. moschata* Duchesne and *C. maxima* Duchesne) and wild *Cucurbita* species. All the *C. pepo* and *C. maxima* accessions behaved as highly susceptible, but four *C. moschata* accessions were highly resistant, two of them after both mechanical and whitefly inoculation, remaining symptomless with a reduced viral accumulation (Sáez et al., 2016).

Genetic resistance to ToLCNDV has also been characterized in some other species belonging to different families. In *Solanum habrochaites* S. Knapp & D.M. Spooner, a wild species related to tomato, three dominant genes are responsible for the resistance (Rai et al., 2013). In *L. cylindrica*, a popular cucurbit vegetable in India, a dominant monogenic resistance was reported (Islam et al., 2010, 2011). More recently, in melon, Sáez et al. (2017) found one major *locus* in chromosome 11 and two additional regions in chromosomes 12 and 2 that control resistance to ToLCNDV. In this context, the purpose of this study was to map the quantitative trait loci (QTL) associated with the resistance to ToLCNDV in *C. moschata* using segregating populations derived from these resistant sources and a susceptible accession of this species, and to confirm this resistance in interspecific

C. moschata × *C. pepo* populations as the first step to transfer the resistance to zucchini.

MATERIALS AND METHODS

Plant Material

In this work, we selected two *Cucurbita moschata* accessions (PI 604506 and PI 381814), previously reported (Sáez et al., 2016) as symptomless or with slight symptoms after whitefly and sap inoculation with ToLCNDV, to study the genetic control of the resistance. PI 604506 is the improved pumpkin cultivar Large Cheese from the United States and PI 381814, an Indian landrace. The Chinese *C. moschata* accession PI 419083 was used as susceptible control. Seeds of the three accessions were first provided by the United States Department of Agriculture-National Plant Germplasm System (USDA-NPGS) genebank, then fixed by selfing and multiplied by the cucurbits breeding group at the Institute for the Conservation and Breeding of Agricultural Biodiversity (COMAV), and stored at the COMAV genebank.

Virus Source and Mechanical Inoculation

To generate the viral inoculum source, susceptible zucchini plants were agroinfiltrated with an infectious clone based on the Spanish isolate of ToLCNDV (99% nucleotide identity with the sequences of the A and B viral genomic particles: KF749224 and KF749225 (Juárez et al., 2014), following the procedure described in Sáez et al. (2016).

The tissue of symptomatic leaves from 15 days post-ToLCNDV agroinoculation plants was crushed in a mortar together with inoculation buffer [50 mM potassium phosphate (pH 8.0), 1% polyvinylpyrrolidone 10, 1% polyethylene glycol 6000, 10 mM 2-mercaptoethanol and 1% activated charcoal] in a 1:4 (*w*/*v*) proportion (López et al., 2015). The homogenate was used to mechanically inoculate all plants at the stage of one true leaf, dusting on the true leaf and on one cotyledon with Carborundum 600 mesh and scratching with a cotton swab dipped in the blend. Inoculated plants were grown in a climatic chamber, and disease progression was monitored. Symptomless plants 15 days after mechanical inoculation (dpi) were reinoculated to avoid escaping to the infection.

Generation of F$_1$ and Segregating Populations

Ten seeds of each *C. moschata* accession were disinfected and germinated as described by Sáez et al. (2016). Seedlings were transplanted to pots and grown in climatic chamber under controlled conditions (photoperiod of 16 h day at 25°C and 8 h night at 18°C and 70% of relative humidity). Subsequently, plants were moved to a greenhouse and crossed to obtain three F$_1$ progenies: F$_1$ PI 419083 × PI 604506, F$_1$ PI 419083 × PI 381814, and F$_1$ PI 604506 × PI 381814. Eight plants of each parent and the corresponding hybrids were mechanically inoculated with ToLCNDV as described above and phenotyped according to symptomatology and viral accumulation as described below.

Eight additional plants of the *C. moschata* parents were cultivated in a greenhouse along with eight plants of the F$_1$ progenies. To generate segregating populations, F$_1$ plants were selfed to obtain F$_2$ progenies and backcrossed to plants of PI 604506, PI 381814, and PI 419083 to generate the BC1$_{PI\ 604506}$, BC1$_{PI\ 381814}$, and BC1$_{PI\ 419083}$ populations, respectively. All these segregating populations were screened against ToLCNDV with the same inoculation and phenotyping methodology, using three plants of each *C. moschata* accession as controls. F$_2$ and BC1 derived from F$_1$ PI 419083 × PI 381814 were obtained later because of the influence of the local climate conditions in PI 381814 vegetative growth, causing late-flowering and slow development of fruits. Hence, we studied first the genetic control of the resistance to ToLCNDV in the segregating populations derived from PI 604506, and results were validated in F$_2$ and BC1 coming from F$_1$ PI 419083 × PI 381814.

Symptoms Evaluation and Quantification of the Viral Accumulation

Symptomatology was evaluated in all plants at 15 and 30 dpi using the visual scale described by López et al. (2015). Symptoms score ranged from 0 (absence of symptoms) to 4 (highly severe symptoms), classifying as resistant those plants with symptoms scored 0 or 1 and as susceptible those with symptoms scored from 2 to 4. The goodness-of-fit between the expected and observed segregation ratios resistant/susceptible plants was analyzed by chi-squared (χ^2) test ($p < 0.05$) in the F$_2$ and BC1-segregating populations.

The relative ToLCNDV accumulation in each plant was determined at 30 dpi by quantitative PCR (qPCR). Total DNA from apical leaves was extracted using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1990) and quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, United States). DNA was diluted with sterile-deionized water to a final concentration of 5 ng μl^{-1}. Three biological replicates were done for each parental genotype, and all plants of the assay were analyzed in three technical replicates using a *LightCycler*® 480 System (Roche). In each qPCR reaction, 15 ng of genomic DNA were used as templates, in a final volume of 15 μl. We used 7.5 μl of 2 × iTaq^{TM} universal SYBR® Green Supermix (BIO-RAD) and 1.5 μl (100 nM) of each primer and 1.5 μl of H$_2$O. Primers ToLCNDVF1 (5′-AATGCCGACT ACACCAAGCAT-3′, positions 1145–1169) and ToLCNDVR1 (5′-GGATCGAGCAGAGAGTGGCG-3′, positions 1399–1418) were used for the amplification of a 273-bp fragment of viral DNA-A. The single-copy gene *CpACS2* was amplified in all samples as internal control using the primers CpACS2F (5′-ACT CGATCAACTTCGAGCAAA-3′) and CpACS2R (5′-GCCTA TCCAAAGACCTCGGCCTTCCC-3′). Both ToLCNDVF1/R1 and *CpACS2* primers were used in previous works by Sáez et al. (2016). Cycling conditions consisted of incubation at 95°C for 5 min, 45 cycles of 95°C for 5 s and 60°C for 30 s. Relative ToLCNDV levels were calculated using the $2^{-\Delta Ct}$ expression, a variation of the Livak method (Livak and Schmittgen, 2001;

Bio-Rad Laboratories, 2006), where ratio (target/reference) = $2^{-\Delta Ct} = 2^{-[Ct(viraltarget) - Ct(reference gen)]}$.

QTL Analysis in *C. moschata* F_2 Population Derived From PI 419083 × PI 604506

PI 604506 and PI 419083 accessions were included in an RNAseq analysis, performed in the frame of a *de novo* assembly of the zucchini genome project (Montero-Pau et al., 2018), and their transcriptome sequences were used to generate the single nucleotide polymorphism (SNP) panel used here. SNPs were selected by aligning each sequence to the version 1 of the *C. moschata* cv. *Rifu* genome (Sun et al., 2017), available at the Cucurbit Genomics Database[1]. We used the Bowtie2 tool with the very-sensitive-local argument. Variant calling was performed using Freebayes version 1.0.2 (Garrison and Marth, 2012), excluding alignments from the analysis if they had a mapping quality < 40, alleles with quality under 20, and filtering SNPs with minimum count of 10. A set of 137 SNPs evenly distributed throughout the *C. moschata* genome (**Supplementary Table 1**) were selected and used to genotype PI 604506, PI 419083, their derived F_1, and 134 plants of the corresponding F_2 population.

All plants were genotyped using the Agena Bioscience iPLEX® Gold MassARRAY (Agena Biosciences) system at the Epigenetic and Genotyping unit of the University of Valencia [Unitat Central d'Investigació en Medicina (UCIM), Faculty of Medicine, Spain]. Total DNA was extracted from the tissue of young leaves, using the protocol described above, and quantified and adjusted to 15 ng μl^{-1}. F_2 genotyping results were run in MAPMAKER 3.0 (Lander et al., 1987; Lincoln et al., 1992) with the Kosambi map function, obtaining the genetic position of each marker.

To identify markers linked to the resistance to ToLCNDV derived from the PI 604506 accession, a QTL analysis was performed using symptoms at 15 and 30 dpi and ToLCNDV relative accumulation at 30 dpi as quantitative traits, and a qualitative score of resistance (0 susceptible phenotype and 1 resistant phenotype) assigned to each plant according to symptoms and viral accumulation. We used the Kolmogorov–Smirnov test to check the normality assumption of traits distribution. Since the traits were not normally distributed, Kruskal–Wallis non-parametric test was used for QTL detection using the MapQTL version 4.1 software (Van Ooijen, 2009), considering as significant associations those with $p < 0.05$. Since $2^{(-\Delta Ct)}$ values have a skewed distribution, we used the original ΔCt data for QTL analysis. The binary qualitative trait of resistance was also analyzed by logistic regression model, with a significance level of $\alpha = 0.05$.

In addition, a composite interval mapping approach (CIM, Zeng, 1994) was applied in Qgene 4.0 (Joehanes and Nelson, 2008), using the genetic map previously generated with this F_2. The logarithm of the odds ratio (LOD) threshold was calculated using a 1,000 permutations test per trait, for $p < 0.05$. The percentage of phenotypic variance explained (R^2), the additive and dominance effects, degree of dominance, and the interval

position of the QTL in accordance with a 2-U LOD drop was estimated for the highest significant peak LOD. *Loci* identified with both methods (Kruskal–Wallis and CIM) were considered true QTLs of putative interest.

Validation of the QTL of Chromosome 8 in Additional *C. moschata* Segregating Populations Derived From PI 419083, PI 604506, and PI 381814

The previous analysis allowed detecting a major QTL responsible for the resistance in chromosome 8. To confirm this QTL in additional *C. moschata*-segregating populations and to introgress the candidate region in chromosome 8 of *C. moschata* in the zucchini (*C. pepo*) background (the cucurbit crop more severely affected by ToLCNDV), a new set of 19 SNPs of the chromosome 8 candidate region was implemented in a new Agena Bioscience platform. These new SNPs were selected to be useful for both purposes. The transcriptomic sequences of PI 604506, PI 381814, and PI 419083 (obtained in the RNAseq analysis by Montero-Pau et al., 2018) were aligned to the *C. pepo* genome (Zucchini accession MU-CU-16), available at the Cucurbit Genomics Database[2], using Bowtie2. Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013) was used to detect variations between sequences, and those polymorphic SNPs between resistant (PI 604506 and PI 381814) and susceptible (PI 419083 and MU-CU-16) genotypes were selected. This Agena platform was employed to genotype a subset of 131 plants of the previously genotyped F_2 (PI 419083 × PI 604506), 121 of F_2 (PI 419083 × PI 381814), 31 BC1$_{PI\ 604506}$, and 73 of BC1$_{PI\ 381814}$.

For further saturation of the candidate region, five additional SNPs, not integrated in the new Agena Bioscience set, were designed with the same requirements and used to genotype the F_2 (PI 419083 × PI 604506) population by high-resolution melting (HRM) (Vossen et al., 2009). PRIMER3 software (Untergasser et al., 2012) was employed to design the oligonucleotides for the HRM analysis. The genomic positions of all these new SNPs (Agena Bioscience platform and HRM markers) and their flanking sequences are shown in **Supplementary Table 2**.

A new map of chromosome 8 was constructed with 24 SNP markers (3 and 16 SNPs from the first and second Agena platforms, respectively, and 5 HRM), using genotyping results of F_2 (PI 419083 × PI 604506). MAPMAKER 3.0 (Lander et al., 1987; Lincoln et al., 1992) software and the Kosambi map function were employed to generate the new map. The genetic distances of the new map were used in a second QTL analysis, with the F_2 (PI 419083 × PI 604506) population, following the same procedure described above. Means of symptom scores at 30 dpi of plants from F_2 (PI 419083 × PI 381814), BC1$_{PI\ 604506}$, and BC1$_{PI\ 381814}$ populations classified according to the marker classes (a, b, and h for F_2 and h and a for BC1) were analyzed by ANOVA and Bonferroni multiple range tests using STATGRAPHIC Centurion XVI.I statistic software, to evaluate

---

[1] http://cucurbitgenomics.org

[2] http://cucurbitgenomics.org

differences between means, considering statistically significant differences when $p \leq 0.01$.

Validation of the QTL in the Interspecific Cross *C. pepo* × *C. moschata*

An interspecific cross between the ToLCNDV-susceptible *C. pepo* accession MU-CU-16 (Sáez et al., 2016) and the resistant *C. moschata* accession PI 604506 provided five F_1 seeds that were germinated as described above. Four seedlings were moved to a greenhouse and selfed to obtain F_2 (MU-CU-16 × PI 604506) generation. The remaining F_1 seedling and 176 plants of F_2 (MU-CU-16 × PI 604506) were screened by mechanical inoculation of ToLCNDV. Symptoms and viral titers were determined by the same procedure described above.

This *Cucurbita*-interspecific F_2 population was genotyped with the new Agena Bioscience platform and the five HRM SNP markers of chromosome 8. The genotyping results were used to construct a new genetic map of chromosome 8 and to perform an additional QTL analysis as described above.

Genomic Variation, Structural Variants, and Synteny

To obtain a more detailed view of the underlying genomic variation in the candidate region, both *C. moschata* resistant and susceptible parents (PI 604506 and PI 419083) were fully sequenced. Raw reads are deposited in the National Center for Biotechnology Information (NCBI) under BioProject PRJNA604046. Genomic DNA was obtained from fresh tissue using CTAB extraction, and a pair-end library (2 × 150 bp) was built for each accession. Libraries were sequenced as part of an Illumina HiSeq 2000 lane by Polar Genomics (Ithaca, NY, United States). Reads were cleaned using the *ngs_crumbs* software[3] to eliminate adapters, low-quality bases (Phred quality < 25 in a 5-bp window), reads shorter than 50 bp, and duplicated sequences. Clean reads were mapped against the reference *C. moschata* genome using *bwa-mem* (Li, 2013), and variant calling was performed using Freebayes version 1.1.0 (Garrison and Marth, 2012) after filtering reads with a mapping quality cutoff MAPQ lower than 57. To study the potential effect of the genetic changes, SNPs were annotated based on its predicted effect on the gene using SNPEff v4.3 (Cingolani et al., 2012). Differences in sequencing genome coverage between both accessions were studied to explore possible genomic deletions. Read coverage along the candidate region was calculated using samtools v.1.9 (Li et al., 2009), and we checked if coverage deviated from the 99% confidence interval of the observed coverage for each accession assuming a log-normal distribution. Confidence interval for the log-normal distribution was calculated using the function *elnorm* of R package "EnvStats" (Millard, 2013). In addition to that, the structural variant caller Manta v.1.6 (Chen et al., 2016) was used to check for differential large insertion/deletions. Identification of putative paralogs of the genes in the candidate region was done with OrthoMCL (Li et al., 2003).

Identification of syntenic regions between *C. moschata* and *C. pepo* and *C. melo* was done by nucleotide basic local alignment search tool (BLAST) of each gene within the candidate region of *C. moschata* against the other two genomes. BLAST hits were filtered using an *E* value cutoff of 10^{-20} and a minimum overlap between sequences of 70%. For *C. pepo*, to inspect for possible insertion/deletions, a dot plot comparing chromosome 17 region of *C. pepo* and chromosome 8 of *C. moschata* was built based on the alignment of both sequences using LAST (Kielbasa et al., 2011). For *C. melo*, the module of Tripal "SyntenyViewer," available in cucurbitgenomics.com, was used to visualize the synteny.

New *C. moschata* and *C. pepo* genome assemblies have become recently available[4,5] (online availability since November 2019), but after finishing the analysis that we showed here. Our results were checked through alignments with the new assemblies to avoid misinformation.

In addition to the analysis of the genomic sequences, SNPs discovered using the available RNAseq data (Montero-Pau et al., 2018) from the three *C. moschata* accessions used as parentals in the previous crosses and six additional *C. moschata* from different origins that exhibited susceptibility to ToLCNDV in previous works (López et al., 2015; Sáez et al., 2016) were also annotated using SNPEff.

RESULTS

Response to ToLCNDV of F_1 Progenies

The inoculation assay showed that the two *C. moschata* accessions resistant to ToLCNDV, PI 604506 and PI 381814, remained totally symptomless or with only slight symptoms (score from 0 to 1) at 30 dpi, contrasting with the severe mosaic developed in the susceptible control (score 4), PI 419083 (**Figure 1**). F_1 plants of the two susceptible × resistant crosses were highly susceptible, displaying a similar symptomatology as PI 419083 at 30 dpi. Conversely, F_1 progeny derived from the cross between the two *C. moschata*-resistant accessions remained symptomless throughout the essay (**Figure 1**).

Strong correlation between symptom severity and viral titers was observed ($r^2 = 0.73$, $p = 0.030$) after measuring relative ToLCNDV accumulation by qPCR. In accordance with their resistant behavior, PI 604506, PI 381814, and the F_1 (PI 604506 × PI 381814) had viral titers, on average, 7.8×10^3 times lower than those of the susceptible control PI 419083 and the two F_1 derived from it (**Figure 1**).

The fact that F_1 progenies derived from the two susceptible × resistant crosses were susceptible, while the F_1 derived from the resistant × resistant cross was resistant, suggests a recessive genetic control of the resistance in both accessions, controlled by common genes. A further analysis of the genetic control of the resistance was performed in F_2 and BC1 populations.
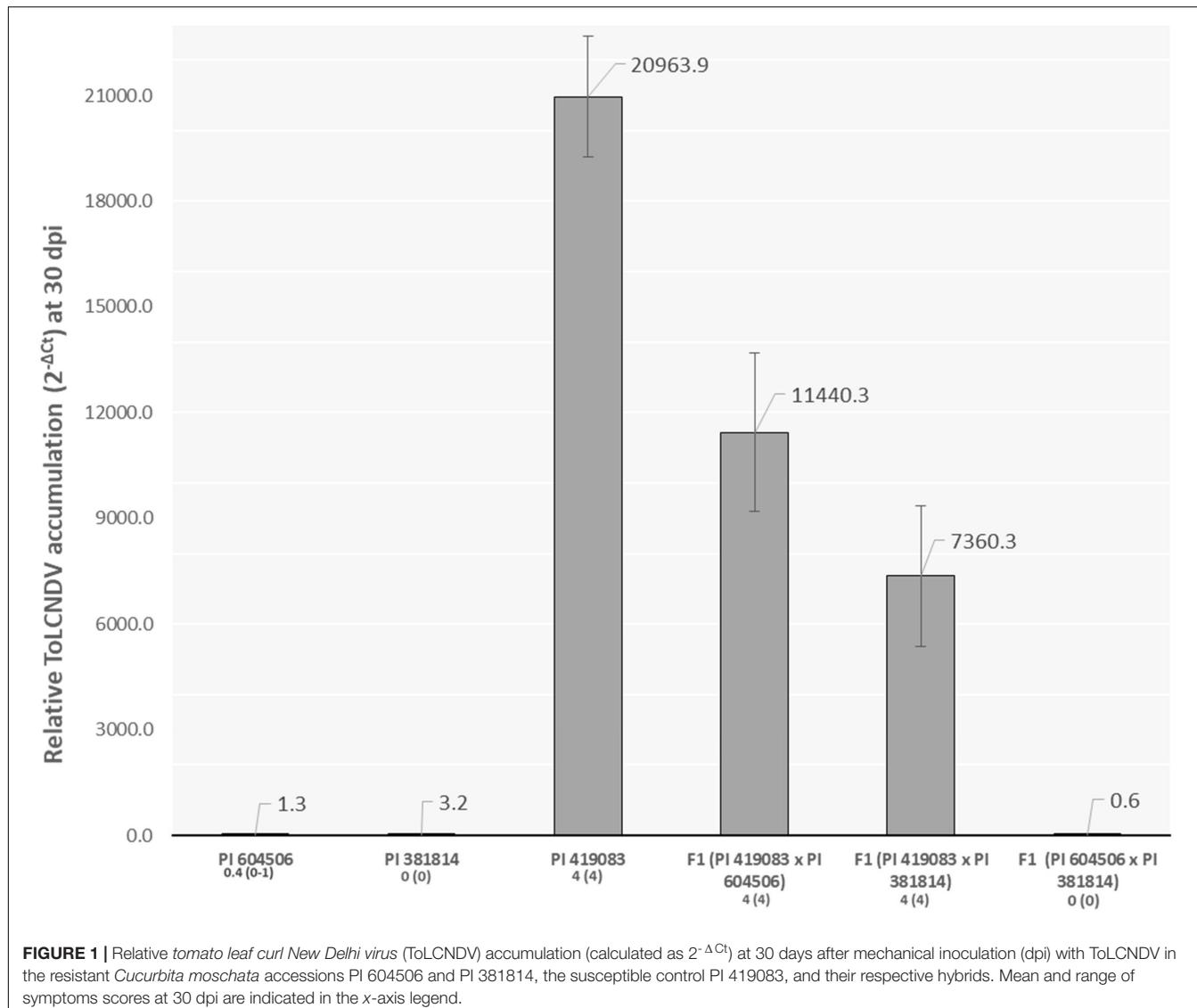
FIGURE 1 | Relative *tomato leaf curl New Delhi virus* (ToLCNDV) accumulation (calculated as $2^{-\Delta Ct}$) at 30 days after mechanical inoculation (dpi) with ToLCNDV in the resistant *Cucurbita moschata* accessions PI 604506 and PI 381814, the susceptible control PI 419083, and their respective hybrids. Mean and range of symptoms scores at 30 dpi are indicated in the *x*-axis legend.

Response to ToLCNDV of Segregating Populations Derived From the Cross Between Resistant and Susceptible *C. moschata* Accessions

F_2 and $BC1_{PI\ 604506}$ populations, derived from the F_1 PI 419083 × PI 604506, segregated for symptoms severity. **Table 1** shows resistant:susceptible plants segregation, according to symptomatology at 30 dpi. At the end of the assay, 38 plants of F_2 remained symptomless (score 0), and 96 showed severe symptomatology (scores 2–4). The X^2 test indicated that this segregation fitted to a 1:3 (resistant:susceptible) ratio expected for a single recessive gene for resistance (p = 0.43) (**Table 1**). To further characterize the response to ToLCNDV, virus accumulation was estimated in the segregating population F_2 (PI 419083 × PI 604506) by qPCR (**Figure 2**). On average, viral titer strongly correlated to symptoms severity following an exponential model (r^2 = 0.82, p = 0.035). All plants developing

mosaic, deformation, or short internodes had high viral titers, whereas in the symptomless plants, ToLCNDV accumulation was detected at very low concentrations. On average, the viral accumulation ($2^{-\Delta Ct}$) in susceptible plants was 2.2×10^3 times higher than in resistant plants. Since viral titer is in concordance with symptoms development, symptom scores were used to phenotype the response to ToLCNDV in plants of the remaining F_2 and BC1 populations. In $BC1_{PI\ 604506}$, 33 plants were resistant (score 0) and 26 were susceptible (scores 2–4). This segregation also fitted to a 1:1 ratio expected for a single recessive gene (p = 0.44) (**Table 1**). In accordance with the occurrence of a single recessive gene controlling the resistance, all plants of the $BC1_{PI\ 419083}$ generation had severe symptoms at the end of the assay.

Symptom segregation ratios observed in the F_2 (PI 419083 × PI 381814) and $BC1_{PI\ 381814}$ populations also fitted to one recessive gene for resistance null hypothesis in X^2 test (**Table 1**). Forty and 43 plants of F_2 and $BC1_{PI\ 381814}$, respectively, remained symptomless (score 0), and 81 and 30

TABLE 1 | Segregation of resistant/susceptible plants in F_2 and BC progenies at 30 days after mechanical inoculation with *tomato leaf curl New Delhi virus* (ToLCNDV).

Female parent	Male parent	Progeny	Resistant: susceptible segregation	Ratio	X^2 test[a]
PI 419083 (S)[b]	PI 604506 (R)[b]	F_2	38:96	1:3	0.6368 (*p* = 0.43)
PI 419083 (S)	PI 604506 (R)	$BC_{PI\ 419083}$	0:90	0:1	−
PI 419083 (S)	PI 604506 (R)	$BC_{PI\ 604506}$	33:26	1:1	0.6102 (*p* = 0.44)
PI 419083 (S)	PI 381814 (R)	F_2	40:81	1:3	3.7713 (*p* = 0.047)
PI 419083 (S)	PI 381814 (R)	$BC_{PI\ 381814}$	43:30	1:1	1.9726 (*p* = 0.16)
PI 604506 (R)	PI 381814 (R)	F_2	160:0	1:0	−

[a]*Probability of the* χ^2 *value calculated for a recessive monogenic expected ratio.* [b]*(S), susceptible genotype; (R), resistant genotype.*
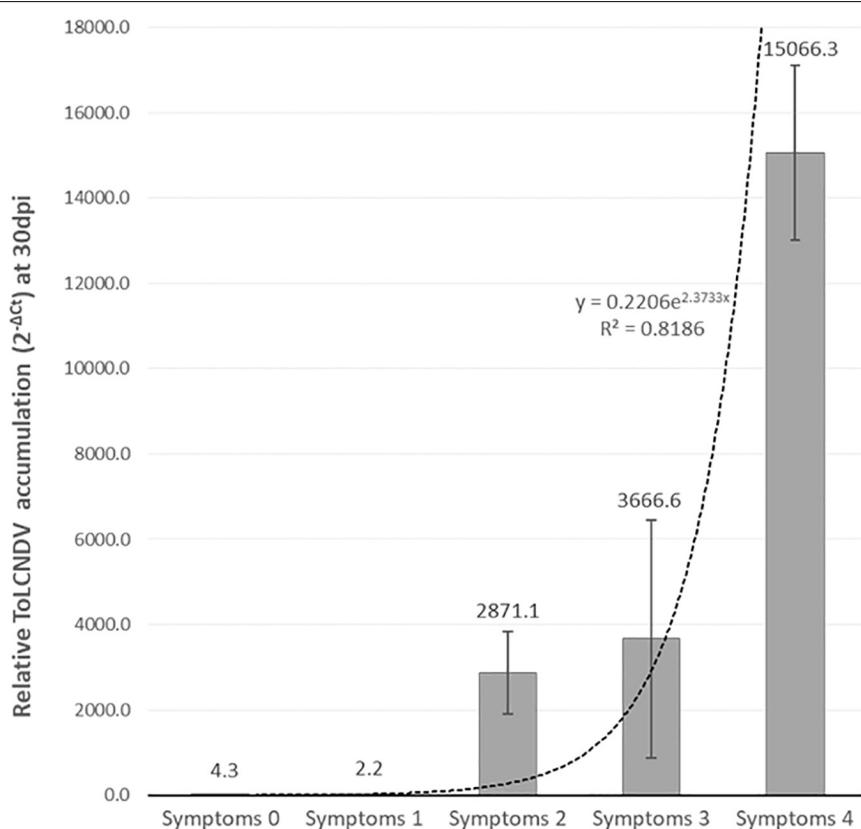
FIGURE 2 | Mean of relative *tomato leaf curl New Delhi virus* (ToLCNDV) accumulation (calculated as $2^{-\Delta Ct}$) in plants within each symptomatic class in F_2 (PI 419083 × PI 604506) at 30 days after mechanical inoculation. Dot line represents the exponential relationship between both variables, which was statistically significant for a confident level of α = 0.05.

plants showed severe symptoms (scores 2–4), with *p* = 0.047 and *p* = 0.16, in both respective populations (**Table 1**).

In accordance with the F_1 results, the 160 plants of the F_2 derived from the resistant × resistant cross PI 604506 × PI 381814 remained totally symptomless along all the assay.

QTL Analysis in F_2 (PI 419083 × PI 604506) Population

The F_2 (PI 419083 × PI 604506) population was genotyped with the 137 SNPs markers evenly distributed throughout *C. moschata* genome and used to construct a linkage map that included 20 linkage groups and spanned a total of 2,681.5 cM of genetic distance, with an average of 22.92 cM between markers (**Supplementary Table 1**). The linkage map was used to identify QTLs involved in ToLCNDV resistance in *C. moschata*, based on genotyping and phenotyping results (symptoms scores at 15 and 30 dpi, virus titer at 30 dpi, and the qualitative resistance score) of F_2 (PI 419083 × PI 604506) population. QTL analysis, performed using non-parametric Kruskal–Wallis test (KW) followed by CIM, resulted in the detection of a major QTL in chromosome 8 (**Table 2**), validated by logistic regression of the

TABLE 2 | Quantitative trait loci (QTLs) identified in the F$_2$ (PI 419083 × PI 604506) segregating population genotyped with the set of 137 single nucleotide polymorphisms (SNPs) evenly distributed through the *C. moschata* genome, using the non-parametric Kruskal–Wallis test and composite interval mapping method.

					Kruskal–Wallis			
Trait	Chr[a]	QTL name	cM (peak position)	Nearest marker[b]	K[c]	Mean a[d]	Mean h[e]	Mean b[f]
Symptoms at 15 dpi	8	ToLCNDVCm_Sy15-8	0	D133	35.51	2.78	2.20	0.46
Symptoms at 30 dpi	8	ToLCNDVCm_Sy30-8	0	D133	55.88	4.00	3.44	1.10
Viral titer at 30 dpi (ΔCt)	8	ToLCNDVCm_VT30-8	0	D133	33.11	−12.76	−10.67	4.24
Resistance (qualitative trait)	8	ToLCNDVCm_Re-8	0	D133	59.14	0.00	0.13	0.73

						CIM (Qgene)				
Trait	Chr[a]	QTL name	cM (peak position)	Interval (cM)[g]	Nearest marker[b]	LOD[h]	Add[i]	Dom[j]	(d/a)[k]	R$^{2(l)}$
Symptoms at 15 dpi	8	ToLCNDVCm_Sy15-8	4.00	0–21.27	D133	10.06	1.52	0.84	0.55	0.29
Symptoms at 30 dpi	8	ToLCNDVCm_Sy30-8	0.00	0–5.43	D133	16.52	1.45	0.90	0.62	0.44
Viral titer at 30 dpi (ΔCt)	8	ToLCNDVCm_VT30-8	0.00	0–6.82	D133	13.97	−4.26	−2.20	0.52	0.39
Resistance (qualitative trait)	8	ToLCNDVCm_Re-8	0.00	0–5.17	D133	17.31	−0.37	−0.23	0.64	0.45

[a]*Chromosome.* [b]*The closest marker to LOD peak.* [c]*K*: the Kruskal–Wallis test statistic, with a significant level of 0.0001.* [d]*Mean of the PI 419083 genetic class in each marker.* [e]*Mean of the PI 419083/PI 604506 genetic class in each marker.* [f]*Mean of the PI 604506 genetic class in each marker.* [g]*Interval position of the putative QTL, identified in the F$_2$ (PI 419083 × PI 604506) in cM on the genetic map according with a LOD drop of 2.* [h]*LOD higher logarithm of the odds score.* [i]*Add additive effect of the PI 419083 allele.* [j]*Dom dominant effect of the PI 419083 allele.* [k]*d/a degree of dominance.* [l]*R^2 percentage of phenotypic variance explained by the QTL.*

qualitative trait of resistance (data not shown). Four QTLs, all located in almost the same genetic position, showed significant association with all the traits evaluated, explaining a proportion between 29.0 and 45.0% of the observed phenotypic variance. All QTLs (ToLCNDVCm_Sy15-8, ToLCNDVCm_Sy30-8, ToLCNDVCm_VT30-8, and ToLCNDVCm_Re-8) were located close to D133 (physical position, 1,366,729 bp), with LOD peaks between 10.06 and 17.31.

Narrowing the Candidate Region in Chromosome 8

To validate the major QTL identified in the previous analysis and to increase marker density in the candidate region, F$_2$ (PI 419083 × PI 604506) population was genotyped with the new Agena Bioscience-HRM SNPs set of chromosome 8. Twenty-one out of the 24 new markers (**Supplementary Table 2**) were polymorphic in this population, despite all of them were selected *in silico* as SNP variants between both parents using the IGV software. Genotyping results were employed to generate a new linkage map in this region (**Supplementary Table 2**), covering 72.5 cM, with an average distance between consecutive markers of 3.15 cM, and two clusters of linked markers at 0 and 11.7 cM genetic positions. The QTL analysis was performed using the new map and the new genotyping results (using one selected marker of each of the two clusters of completely linked SNPs) (**Table 3**). ToLCNDVCm_Sy15-8 QTL, associated to the variation of symptoms at 15 dpi, was identified again with both non-parametric Kruskal–Wallis and CIM analysis, near D133 (located at 18.8 cM in this new map) and with similar explained variance, LOD peak, additive, and dominant effects. However, ToLCNDVCm_Sy30-8, ToLCNDVCm_VT30-8, and ToLCNDVCm_Re-8 QTLs, corresponding to traits measured

at the end of the assay (30 dpi), when differences are clearer between resistant and susceptible plants, were closely linked to a new marker with both analysis methods. The closest markers (those linked at 11.7 cM DPM37, DMP39, DMP11, DMP10, DMP42, DMP43, DMP44, DMP41, and snp_8202510 markers) are included in the interval position of the same QTLs identified with Kruskal–Wallis and CIM (between 8 and 14 cM) (**Figure 3A**) and validated with logistic regression of the qualitative trait of resistance, according to their physical and genetic position. The interval of the four QTLs was flanked by DPM34 and D133 markers, with physical positions 561,788 and 1,366,729 bp, respectively. After this further QTL analysis of chromosome 8, the proportion of explained variance was increased, with percentages between 29.5 and 66.0% of R^2.

Validation of the Major QTL in Chromosome 8 in BC1$_{PI\ 604506}$, F$_2$ (PI 419083 × PI 381814) and BC1$_{PI\ 381814}$ Segregating Populations

The Agena Bioscience SNPs panel of chromosome 8 was used to genotype the BC1$_{PI\ 604506}$ derived from the PI 419083 × PI 604506 cross, and the F$_2$ and BC1$_{PI\ 381814}$ populations derived from the PI 419083 × PI 381814 cross. Mean of symptoms scores at 30 dpi were calculated for each genotypic class of selected SNPs located within the defined QTL interval (DMP35 and DMP39) and compared in **Figures 3B–D**. The lowest level of symptoms was observed when plants in the three populations had the PI 604506 or PI 381814 homozygous genotype (b), in DPM35 or DPM39 indistinctly. Plants heterozygous (h) or PI 419083 homozygous (a) in both markers displayed significantly more severe symptomatology.

TABLE 3 | Quantitative trait loci (QTLs) identified in the F$_2$ (PI419083 × PI604506) segregating population genotyped with markers of chromosome 8 of *C. moschata*, using the non-parametric Kruskal–Wallis test and composite interval mapping (CIM).

				Kruskal-Wallis				
Trait	Chr[a]	QTL name	cM (peak position)	Nearest marker[b]	K[c]	Mean a[d]	Mean h[e]	Mean b[f]
Symptoms 15 dpi	8	ToLCNDVCm_Sy15-8	18.8	D133	35.58	2.84	2.22	0.46
Symptoms 30 dpi	8	ToLCNDVCm_Sy30-8	11.7	DMP39[m]	82.96	4.00	3.60	0.46
Viral titer at 30 dpi (ΔCt)	8	ToLCNDVCm_VT30-8	11.7	DMP39[m]	50.38	−12.63	−11.14	−2.42
Resistance (qualitative trait)	8	ToLCNDVCm_Re-8	11.7	DMP39[m]	86.71	0.00	0.10	0.89

m: markers with same genetic position and significance in the analysis: DPM37, DMP39, DMP11, DMP10, DMP42, DMP43, DMP44, DMP41

					CIM (Qgene)				
Trait	Chr[a]	QTL name	cM (peak position)	Interval (cM)[g]	Nearest marker[b]	LOD[h]	Add[i]	Dom[j]	[d/a][k]	R^2[l]
Symptoms 15 dpi	8	ToLCNDVCm_Sy15-8	18.00	4.53–24.00	D133	9.96	1.25	0.60	0.48	0.30
Symptoms 30 dpi	8	ToLCNDVCm_Sy30-8	12.00	8.01–13.46	DMP39	29.76	1.79	1.38	0.77	0.65
Viral titer at 30 dpi (ΔCt)	8	ToLCNDVCm_VT30-8	12.00	10.85–14.15	DMP39	24.37	−5.18	−3.64	0.70	0.59
Resistance (qualitative trait)	8	ToLCNDVCm_Re-8	12.00	8.13–13.83	DMP39	30.92	−0.45	−0.35	0.76	0.66

Genetic positions are according with the new C. moschata × C. moschata linkage map of chromosone 8. [a]Chromosome. [b]The closest marker to LOD peak. [c]K*: the Kruskal–Wallis test statistic, with a significant level of 0.0001. [d]Mean of the PI 419083 genetic class in each marker. [e]Mean of the PI 419083/PI 604506 genetic class in each marker. [f]Mean of the PI 604506 genetic class in each marker. [g]Interval position of the putative QTL, identified in the F$_2$ (PI 419083 × PI 604506) in cM on the genetic map according with a LOD drop of 2. [h]LOD, higher logarithm of the odds score. [i]Add, additive effect of the PI 419083 allele. [j]Dom, dominant effect of the PI 419083 allele. [k]d/a, degree of dominance. [l]R^2 percentage of phenotypic variance explained by the QTL.

QTL Analysis and Validation of the Candidate Region in *C. pepo*

Consistently with the results obtained in F$_1$ from susceptible × resistant *C. moschata* crosses, severe symptoms were developed by F$_1$ *C. pepo* MU-CU-16 × *C. moschata* PI 604506 plants at 15 and 30 dpi (**Figure 4**). This result supports that resistance in PI 604506 has a recessive genetic control. F$_2$ (MU-CU-16 × PI 604506) plants segregated for symptomatology and viral accumulation. Symptoms, including upward and downward curling and severe mosaic of young leaves, short internodes, and bad distorted development, were observed in 124 and 151 F$_2$ (MU-CU-16 × PI 604506) plants at 15 and 30 dpi, respectively. The number of resistant plants decreased from 52 to 25 between 15 and 30 dpi. Nine plants had bad development or died in the course of the infection. On average, virus titers determined by qPCR at 30 dpi were in concordance with symptoms development, with mean of relative viral accumulation expressed as $2^{(-\Delta Ct)}$ of 1.04 ± 0.31 and 49,571.67 ± 9,670.31 in resistant and susceptible plants, respectively. The observed segregation proportion was adjusted to the expected ratio resistant/susceptible plants, in case of one recessive gene responsible on the genetic control of resistance to ToLCNDV at 15 dpi (X^2 = 2.1894, p = 0.14), but not at 30 dpi (X^2 = 10.312, p = 0.0014).

The genetic map of chromosome 8 generated with the genotyping results of the Agena Bioscience-HRM SNPs in the F$_2$ (MU-CU-16 × PI 604506) gave a total genetic length of 21.4 cM, with an average genetic distance between successive markers of 0.98 cM (**Supplementary Table 2**).

The QTL analysis performed in this population show that the QTLs identified in the *C. moschata* populations were stable in the cross with the *C. pepo* accession MU-CU-16. ToLCNDVCm_Sy15-8, ToLCNDVCm_Sy30-8, ToLCNDVCm_VT30-8, and ToLCNDVCm_Re-8 were located in the same region that in *C. moschata* (**Table 4**), physically mapped in chromosome 17. The highest R^2 value (65%) was explained by the ToLCNDVCm_Sy15-8, associated to DMP39 as the nearest marker to the peak LOD. R^2 values were lower in QTLs related to advanced stages of the ToLCNDV infection, mainly in the viral titer at 30 dpi (ΔCt) trait. In these cases, the nearest markers to the LOD peaks were DMP39 and snp_7926165 in ToLCNDVCm_Sy30-8 (Kruskal–Wallis and CIM tests, respectively), snp_7926165 in ToLCNDVCm_VT30-8 and DMP35, and snp_7926165 in ToLCNDVCm_Re-8 (Kruskal–Wallis and CIM tests, respectively). Logistic regression validate the occurrence of ToLCNDVCm_Re-8 QTL. According to the 2-LOD drop confidence intervals, the position interval where the four QTLs are comapping in chromosome 17 of *C. pepo* genome (v.4.1) is delimited between DMP34 (7,658,175 bp) and DMP41 (8,165,929).

After both QTL analysis of chromosome 8, a consensus candidate region considered as responsible for ToLCNDV resistance in *C. moschata*, was established between DMP34 (561,788) and snp_8202510 (1,116,660).

Genomic Variation, Structural Variants, and Synteny

The alignment between the reference assemblies of *C. moschata* and *C. pepo* used for mapping purposes in the current paper[6]

---

[6]http://cucurbitgenomics.org
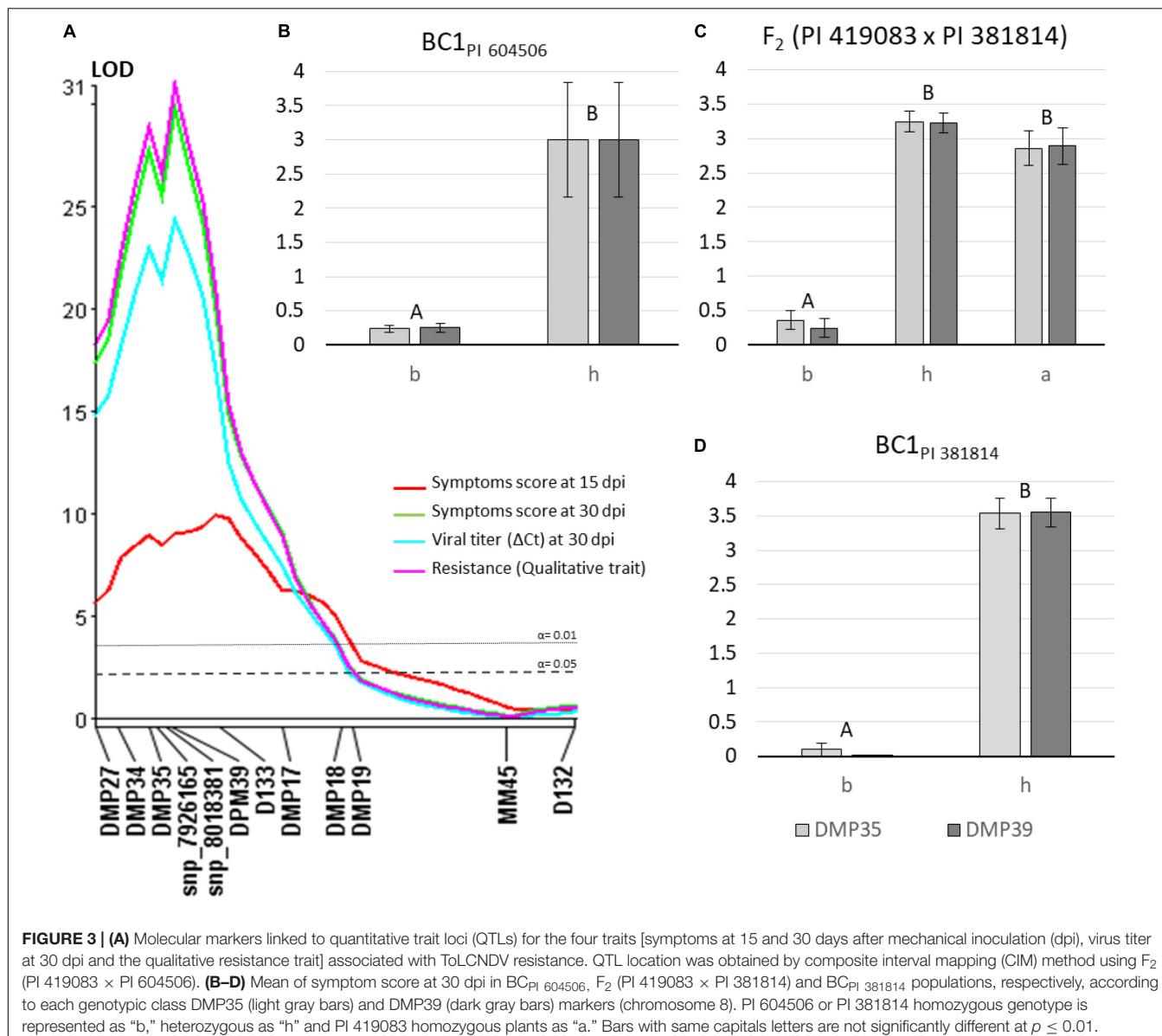
FIGURE 3 | (A) Molecular markers linked to quantitative trait loci (QTLs) for the four traits [symptoms at 15 and 30 days after mechanical inoculation (dpi), virus titer at 30 dpi and the qualitative resistance trait] associated with ToLCNDV resistance. QTL location was obtained by composite interval mapping (CIM) method using F$_2$ (PI 419083 × PI 604506). **(B–D)** Mean of symptom score at 30 dpi in BC$_{PI\ 604506}$, F$_2$ (PI 419083 × PI 381814) and BC$_{PI\ 381814}$ populations, respectively, according to each genotypic class DMP35 (light gray bars) and DMP39 (dark gray bars) markers (chromosome 8). PI 604506 or PI 381814 homozygous genotype is represented as "b," heterozygous as "h" and PI 419083 homozygous plants as "a." Bars with same capitals letters are not significantly different at $p \leq 0.01$.

and the new assemblies available in November 2019[7] showed no significant effect on the QTL region studied here (**Supplementary Figure 1**). Consequently, we keep working with the previous reference versions of both genomes.

A total of 53.2 and 31.5 million genomic clean reads were obtained from PI 604506 and PI 419083, respectively, and approximately more than 97% of them mapped against the *C. moschata* v.1 reference genome. No large structural variants were found between both accessions, and the read genome coverage was similar among them (**Figure 5**), which indicates that there are no deletions causing the observed phenotype. Some genomic positions show significant deviations for the expected coverage in both accessions (**Figure 5**), which could indicate some assembly errors on the reference genome.

After filtering for mapping quality, 28.2 and 18.6 million reads were kept. A total of 1,220,940 SNPs were found to be variable between both parental accessions, and 2,748 were located in the candidate region in chromosome 8. Out of them, nine SNPs had a predicted high impact (either a frameshift or missense variant, a stop codon gain/loss, or a splice site variant) and located within six genes (**Supplementary Table 3**). Two of these markers are located in the same genes where SNPs used in mapping (snp_7926165 and DMP44) were detected to be linked to ToLCNDV resistance [CmoCh08G001470 encoding a BZIP transcription factor bZIP80 (835,327 to 841,749 bp) and CmoCh08G001770 encoding an unknown protein (1,047,526– 1,051,835 bp)]. The remaining seven SNPs with predicted high impact were located in three additional genes of this interval [CmoCh08G001130 encoding a Ribosome inactivating protein (583,200–588,238 bp), CmoCh08G001780 encoding

---
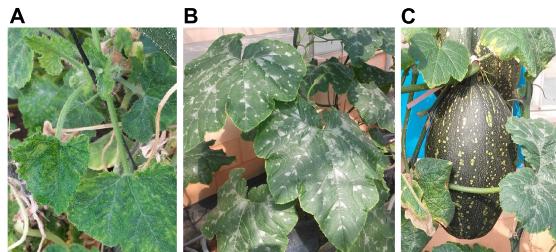
FIGURE 4 | Plants of the interspecific F$_1$ resulting from the cross *C. pepo* MU-CU-16 × *C. moschata* PI 604506. **(A)** Typical symptoms of *tomato leaf curl New Delhi virus* including curling, severe mosaic, and short internodes. **(B)** Healthy plant used to obtain F$_2$ progeny by selfing in a greenhouse. **(C)** Detail of F$_1$ (MU-CU-16 × PI 604506) fruit obtained by selfing.

a putative transmembrane protein (1,051,479–1,053,847 bp) and CmoCh08G001880 coding a IQ-DOMAIN 14-like protein (1,097,864–1,102,974)]. In addition, some other SNPs with low, moderate, or unknown modifying effect are placed in genes related to plant virus resistance (**Supplementary Table 3**).

In addition to the genomic SNPs, the transcriptomic sequences of the three parentals and the six additional susceptible *C. moschata* accessions provided 731 SNPs in the candidate region, 94 of them were fixed for different alleles in the PI 604506-resistant accession and in the seven susceptible accessions (**Supplementary Table 3**). PI 381814 transcriptomic sequence had a low coverage in the candidate region, and it was not possible to identify common polymorphisms between the two resistant accessions, PI 604506 and PI 381814. Three SNPs were detected with high predicted effect, all of them were common to those found in the genomic sequences analyzed

and were located in three genes (**Supplementary Table 3**) (CmoCh08G001130 encoding a ribosome-inactivating protein, CmoCh08G001470 encoding a BZIP transcription factor bZIP80 and, CmoCh08G001770 encoding an unknown protein).

The structure of the candidate region was studied in more detail. A whole genome duplication likely occurred in the species that originated the *Cucurbita* genus (Montero-Pau et al., 2018). In fact, the search for putative paralogs of the genes in the chromosome 8 region indicated that 68 out of the 86 genes in the chromosome 8 candidate region could be assigned to an orthogroup, and 58 of them presented at least one paralog gene. These paralog genes are widespread along the genome (**Figure 6**), although it seems that there is a conserved duplicated region of chromosome 8 on chromosome 17. Interestingly, some genes of the candidate region have been identified as single copy in chromosome 8 (**Supplementary Table 4**), without paralog genes in other chromosomes, which is consistent with a major QTL responsible of ToLCNDV resistance.

We also studied the synteny of this region with the susceptible *C. pepo*, which is phylogenetically closely related to *C. moschata*. BLAST alignment showed synteny between chromosome 8 region and chromosome 17 from 7,658,023 to 8,205,474 bp of *C. pepo* (see **Figure 7** and **Supplementary Table 4**). Gene order and orientation is preserved for most genes, but there is one region showing INDELs. Interestingly, the region with a major insertion in *C. pepo*, from 8,108,962 to 8,113,419 bp, is the region in which the MAD-box transcription factor CmoCh08G001760 maps. This region correspond to position 1,024,011 bp of *C. moschata*, located between the 5′ untranslated region (UTR) and the first exonic region of this gene. Specific analysis of this *C. pepo* insertion sequence allowed to detect a long terminal repeats (LTR) retrotransposon of Ty1-copia Retrofit/Ale kind,

TABLE 4 | Quantitative trait loci (QTLs) identified in the F$_2$ (MU-CU-16 × PI 604506) segregating population genotyped with markers evenly distributed in chromosome 8 of *C. moschata*, using the genetic map obtained with this population, using the non-parametric Kruskal–Wallis test and composite interval mapping (CIM).

			Kruskal–Wallis					
Trait	Chr[a]	QTL name	cM (peak position)	Nearest marker[b]	K[c]	Mean a[d]	Mean h[e]	Mean b[f]
Symptoms 15 dpi	17	ToLCNDVCm_Sy15-8	10.5	DMP39	109.65	3.78	3.49	0.25
Symptoms 30 dpi	17	ToLCNDVCm_Sy30-8	10.5	DMP39	75.99	3.87	3.84	1.73
Viral titer at 30 dpi (ΔCt)	17	ToLCNDVCm_VT30-8	8.3	snp_7926165	28.89	−14.51	−13.83	−6.02
Resistance (qualitative)	17	ToLCNDVCm_Re-8	7.2	DMP35	63.18	0.02	0.02	0.53

			CIM (Qgene)						
Trait	Chr[a]	QTL name	cM (peak position)	Interval (cM)[g]	Nearest marker[b]	LOD[h]	Add[i]	Dom[j]	(d/a)[k]	R^{2}[l]
Symptoms 15 dpi	17	ToLCNDVCm_Sy15-8	10	8.22–10.70	DMP39	40.57	1.76	1.49	0.84	0.65
Symptoms 30 dpi	17	ToLCNDVCm_Sy30-8	8	6.98–11.32	snp_7926165	20.70	1.08	1.05	0.97	0.42
Viral titer at 30 dpi (ΔCt)	17	ToLCNDVCm_VT30-8	8	6.57–11.62	snp_7926165	15.32	−4.25	−3.56	0.84	0.33
Resistance (qualitative)	17	ToLCNDVCm_Re-8	8	6.92–11.33	snp_7926165	17.28	−0.25	−0.25	1.00	0.37

[a]*Chromosome.* [b]*The closest marker to LOD peak.* [c]*K*, the Kruskal–Wallis test statistic, with a significant level of 0.0001.* [d]*Mean of the MU-CU-16 genetic class in each marker.* [e]*Mean of the MU-CU-16/PI 604506 genetic class in each marker.* [f]*Mean of the PI 604506 genetic class in each marker.* [g]*Interval position of the putative QTL, identified in the F$_2$ (MU-CU-16 × PI 604506) in cM on the genetic map according with a LOD drop of 2.* [h]*LOD higher logarithm of the odds score.* [i]*Add additive effect of the MU-CU-16 allele.* [j]*Dom dominant effect of the MU-CU-16 allele.* [k]*d/a degree of dominance.* [l]*R^2 percentage of phenotypic variance explained by the QTL.*
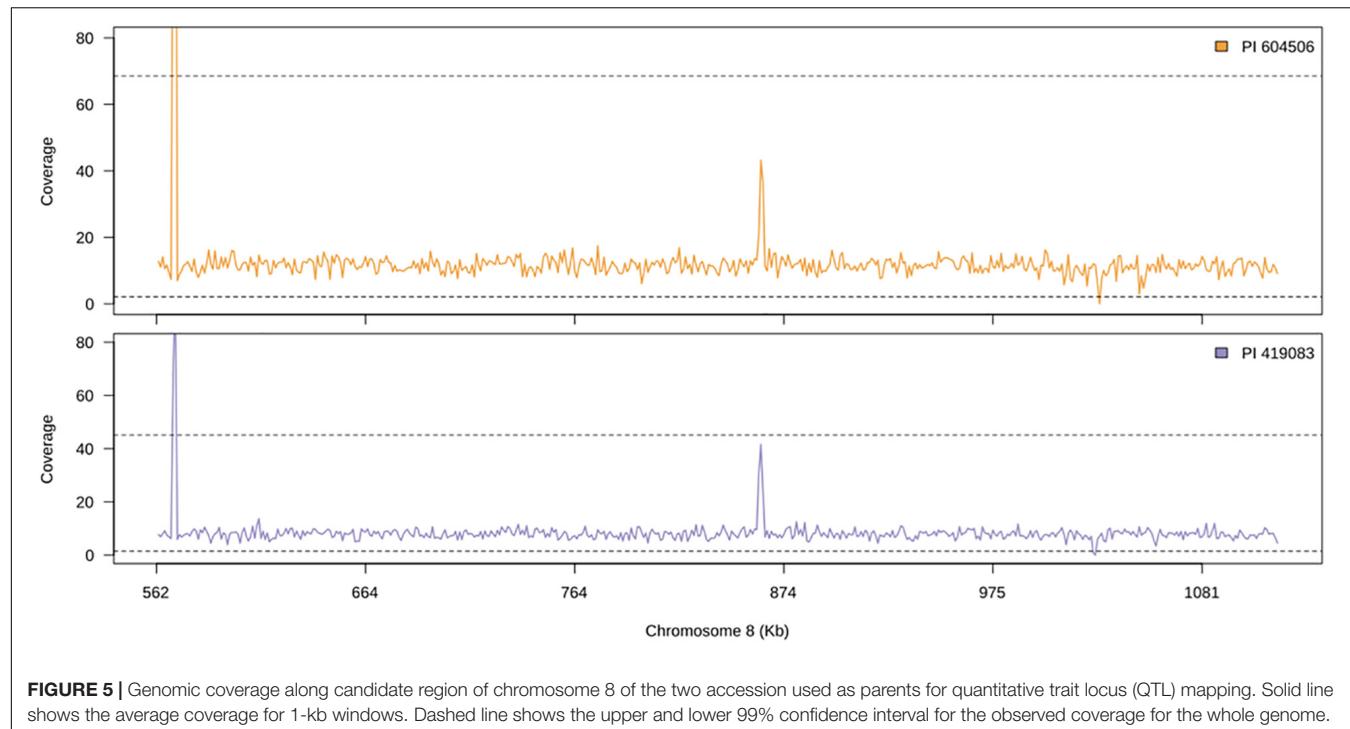
FIGURE 5 | Genomic coverage along candidate region of chromosome 8 of the two accession used as parents for quantitative trait locus (QTL) mapping. Solid line shows the average coverage for 1-kb windows. Dashed line shows the upper and lower 99% confidence interval for the observed coverage for the whole genome.
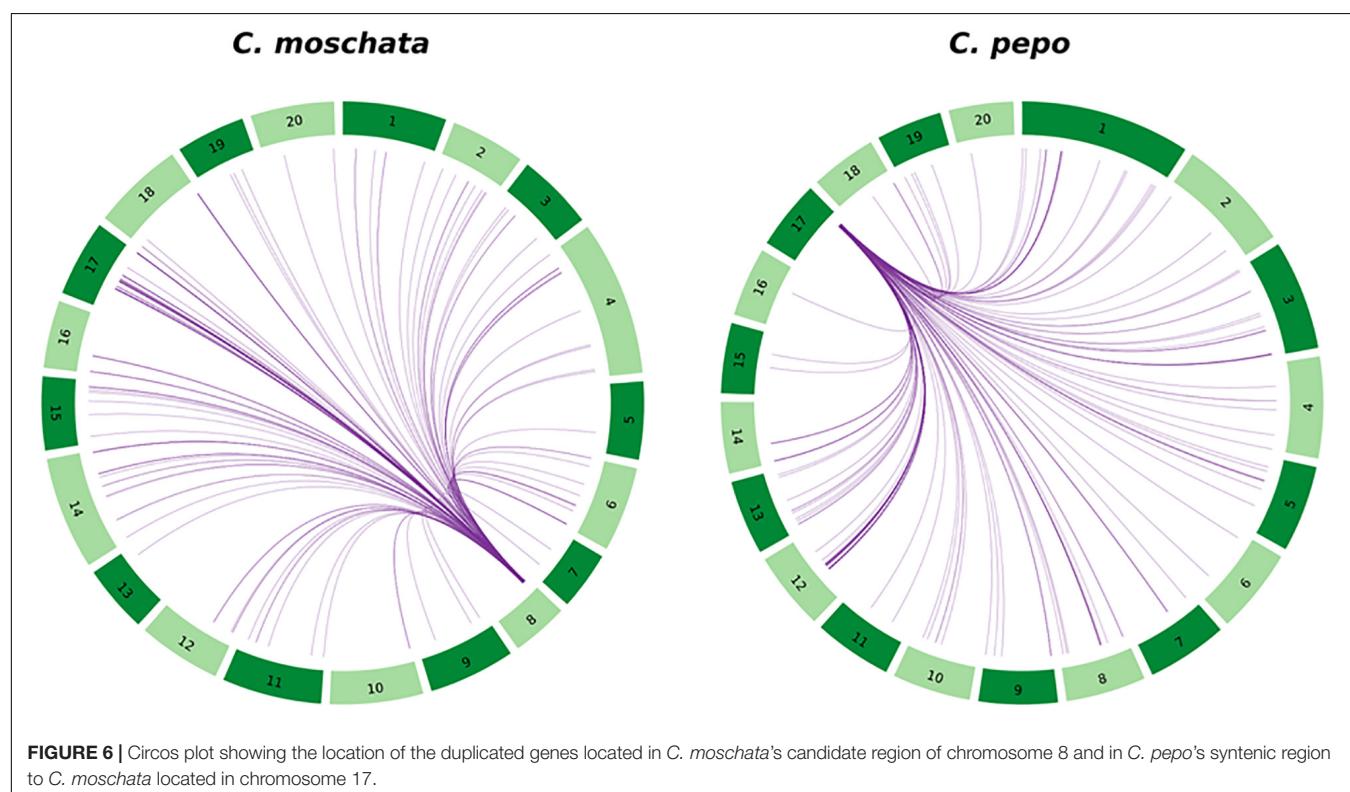
FIGURE 6 | Circos plot showing the location of the duplicated genes located in *C. moschata*'s candidate region of chromosome 8 and in *C. pepo*'s syntenic region to *C. moschata* located in chromosome 17.

of 3,692 bp length located from 8,109,186 to 8,113,548 bp. This transposable sequence was previously annotated using the annotation procedure for repetitive sequences described in Montero-Pau et al. (2018). **Supplementary Table 6** shows

the annotation results and the fasta sequence of the region. Although this insertion is absent in both resistant (PI 604506) and susceptible (PI 419083) *C. moschata* accessions (**Figure 5**), many polymorphic SNPs between them are located in this

FIGURE 7 | (A) Dot plot showing the alignment between the chromosome 8 of *C. moschata* assembly v.1 and chromosome 17 of *C. pepo* v.4.1. **(B)** Expanded syntenic region where large INDELs have been detected. Blue and red arrows points genes sense.

FIGURE 8 | Synteny between the interval regions of quantitative trait loci (QTLs) detected in chromosome 8 of *C. moschata* and the major QTL in chromosome 11 of *C. melo.* **(A)** Circular representation showing synteny with genome of melon DHL92 (v3.6.1). **(B,C)** Synteny blocks where QTLs linked to *tomato leaf curl New Delhi virus* (ToLCNDV) resistance are located (coded as cmomedB906 and cmomedB910 in the database).

gene, including 5′ UTR and 3′ UTR variants (1,023,872 and 1,047,775 bp), and a missense variant with moderate effect (1,043,369).

BLAST search of the *C. moschata* QTL region against *C. melo* found several syntenic regions. In the case of *C. melo*, highly significant alignments were obtained against chromosome 11 where a major QTL associated with resistance to ToLCNDV is located (Sáez et al., 2017). Results show inversions in SNPs positions between both

species, with at least two points of inversion events and loss of information regions (**Supplementary Table 2**). This syntenic relationship was confirmed with the information displayed by the SyntenyViewer of cucurbitgenomics.org tool. Using chromosome 8 of *C. moschata* as query genome and location, circular representation showed regions of synteny with eight chromosomes of melon DHL92 (v3.6.1), including the candidate region of chromosome 11 (**Figure 8A**). **Figures 8B,C** show the syntenic blocks where ToLCNDV resistance-linked QTLs

are located (coded as cmomedB906 and cmomedB910 in the database), the genomic position covered, and the graphic synteny relationship in both blocks. Furthermore, statistical significance of synteny between homologous genes in the candidate region of *C. moschata* and *C. melo* is presented in **Supplementary Table 5**. Seventeen genes are shared by both candidate regions, including the MAD-box transcription factor CmoCh08G001760 and the transmembrane protein CmoCh08G001780 where INDELs or high-effect SNPs have been identified.

DISCUSSION

In this work, we evaluated the resistance to ToLCNDV previously described in the two *C. moschata* accessions PI 604506 and PI 381814 using mechanical inoculation (Sáez et al., 2016). Our results confirmed that both genotypes remain symptomless after inoculation assays. The Large Cheese improved cultivar PI 604506 originated in the United States (Burpee Company). Even though the primary center of *C. moschata* diversity is located in Northern South America and Central America, it spread soon to Mexico and later to the Caribbean area and the United States, where it diversified (Decker-Walters and Walters, 2000). The landrace PI 381814 was collected in India, a secondary center of *C. moschata* variation, where resistance to ToLCNDV was found in melon accessions (López et al., 2015). This fact can be related with the coevolution of host and pathogen in this area, in which ToLCNDV was detected for the first time infecting cucurbits many years ago. Indian cucurbits germplasm has been previously used as source of resistances to viral and fungal pathogens (Dhillon et al., 2012; McCreight et al., 2017). Mendelian analysis of symptom segregation in F_2 and BC_1s populations derived from PI 604506 and PI 381814, as well as QTL results, suggested the presence of a major recessive gene in chromosome 8 of *C. moschata* controlling symptoms development and virus titer. Allelism test results, which show resistance in all plants of F_2 (PI 604506 × PI 381814), suggests that alleles of the same *locus* control ToLCDV resistance in both accessions.

The occurrence of a major gene controlling ToLCNDV resistance derived from *C. moschata* sources is consistent with the existence of a major QTL reported to control the resistance to ToLCNDV in melon, derived from the wild Indian accession of *Cucumis melo* subsp. *agrestis* WM-7 (Sáez et al., 2017). Resistance to whitefly transmission of ToLCNDV in sponge gourd (*L. cylindrica*), a cucurbit crop widely cultivated in India (Islam et al., 2010), has also been described to be regulated by a main dominant gene, for which two linked sequence-related amplified polymorphism (SRAP) markers were reported (Islam et al., 2011).

Even though the major QTL linked to the resistance in *C. moschata* was stable in the *C. pepo* × *C. moschata* interspecific progeny, the mendelian segregation of symptoms only fitted to one recessive gene at 15 dpi. The effect of additional minor genes contributing to ToLCNDV resistance that are segregating in this interspecific population could account for these differences. In fact, in melon, besides the major QTL of chromosome 11, two additional minor regions in chromosomes 12 and 2 modifying

the resistant response were identified (Sáez et al., 2017). In a recent publication (Romay et al., 2019), one recessive (*bgm-1*) and two dominant (*Bgm-2* and *Tolcndv*) genes were also found controlling resistance to ToLCNDV in the same Indian accession WM-7. A similar oligogenic control, three dominant genes, has been reported in *S. habrochaites* S. Knapp & D. M. Spooner, a wild species related to tomato, after ToLCNDV agroinoculation (Rai et al., 2013).

The role of the genetic background in resistance to plant viruses is considered determinant in breeding programs when transferring QTLs from one species to another. Gallois et al. (2018) have studied and reviewed the effect of epistatic relationship with QTL analysis on virus resistance, suggesting that a major-effect QTL (proportion of phenotypic variance explained by the QTL $R^2 > 0.60$) could be more susceptible to genetic background influence than minor-effect QTLs. This statement supports the incomplete penetrance obtained when we tried to transfer the QTL conferring resistance to ToLCNDV from chromosome 8 of *C. moschata* into *C. pepo* background. In this work, R^2 percentages of QTLs detected in the F_2 (PI 419083 × PI 604506) at 30 dpi ranged between 53 and 64% (**Table 3**), while in F_2 (MU-CU-16 × PI 604506), the R^2 percentages of QTLs linked to the same candidate region decreased from 15 dpi ($R^2 = 65\%$) to 30 dpi ($R^2 = 33–42\%$). These results suggest the requirement of other *loci* fixed in the *C. moschata* genetic background needed in the mechanism of resistance to ToLCNDV, but segregating in *C. pepo*. With this information, it is recommended to select for resistance at 30 dpi in populations coming from interspecific crosses, as it is the final stage of infection that better reflect the final response of the plants to the virus.

Resistance to ZYMV was found in different *C. moschata* accessions (Munger and Provvidenti, 1987; Paris et al., 1988; Wessel-Beaver, 2005). The resistance in the Portuguese *C. moschata* accession, Menina, is conferred by one dominant gene, Zym-1, in the cross with the susceptible *C. moschata* Waltham Butternut (Paris et al., 1988). However, when the resistance from Menina was introgressed into the *C. pepo,* segregation did not adjust to a single-gene ratio, and other additional dominant genes, Zym-2 and Zym-3, seemed to be involved in the resistance (Paris and Cohen, 2000). According to Pachner et al. (2011), even Zym-1, Zym-2, and Zym-3 together in *C. pepo* do not confer the same level of resistance seen in "Menina." Studies of inheritance of ZYMV resistance showed that the presence of Zym-1 is essential, but must be combined with other six genes to obtain different levels of expression and durability of resistance in *C. pepo* (Pachner et al., 2015; Capuozzo et al., 2017). In accordance with these works, future QTL analysis of F_2 (MU-CU-16 × PI 604506), including genotyping with SNPs covering the whole *Cucurbita* genome, are crucial to reveal epistatic effects of other *loci* affecting ToLCNDV resistance.

The major *locus* for resistance to ToLCNDV in chromosome 8 of both *C. moschata* sources, PI 604506 and PI 381814, is recessively inherited. Recessive resistance genes, or susceptibility genes, because their presence conditions virus susceptibility (Garcia-Ruiz, 2018), are a common defense strategy against plant viruses (Diaz-Pendón et al., 2004; Kang et al., 2005). In cucurbits,

recessive resistance genes have been reported in several viruses. Translation initiation factors eIF(iso)4E and eIF4G confer recessive resistance against a subset of viruses in several crop species (Hashimoto et al., 2016). The *nsv* recessive gene, encoding an eIF4E factor, confers resistance to *melon necrotic spot virus* (Nieto et al., 2006), preventing the accumulation of viral RNA at the single-cell level (Díaz et al., 2002). In potyvirus-infected *Nicotiana benthamiana* leaf tissues, DEAD-box RNA helicase RH8, which share sequence homology with eIF4A, a component of the eIF4F multiprotein complex, is involved in viral genome translation and replication (Huang et al., 2010). We searched for putative eIF4E and eIF4F at the candidate region for ToLCNDV resistance of *C. moschata* annotation reference genome and found that two genes (CmoCh08G001290 and CmoCh08G001490) encoding an ATP-dependent RNA helicase and chromodomain-helicase-DNA-binding protein 1-like protein, respectively, mapped on the candidate region of chromosome 8. Concretely, CmoCh08G001490 is a single-copy gene in *C. moschata*, with a 3′ UTR SNP variant in PI 604506 sequence and is syntenic with a basic leucine zipper (BZIP) domain class transcription factor gene (MELO3C022278) of the chromosome 11 of *C. melo*.

In addition, other strategies have been reported for recessive resistance against viruses. The recessive *cmv1* gene that confers resistance to *cucumber mosaic virus* in melon encodes a vacuolar protein sorting 41 (VPS41) (Giner et al., 2017) involved in membrane trafficking to the vacuole. Membrane components are key factors required for plant infection success, and viral replication is associated with host intracellular membranes (Nicaise, 2014). In the case of *tom1* and *tom2A Arabidopsis* mutants, *tobacco mosaic virus* (TMV) accumulation is suppressed in single cells. Both genes encode transmembrane proteins localized in the tonoplast that are required for tobamovirus replication (Ishibashi et al., 2012). Among the annotated genes within the *C. moschata* candidate region here identified, several genes are related with membrane components. CmoCh08G001420 encodes a vesicle transport protein and CmoCh08G001500 an autophagy-related protein 3. Interestingly, two of the genes where high-impact SNPs have been detected are annotated as putative transmembrane protein (CmoCh08G001780 and CmoCh08G001790), included in the syntenic region between both candidate regions with resistance to ToLCNDV in *C. moschata* and *C. melo*.

Comparative physical mapping revealed a high level of synteny between the candidate regions with the major QTLs controlling ToLCNDV resistance of chromosomes 8 and 11 of *C. moschata* and *C. melo*, respectively (Sáez et al., 2017). The interval of ∼118 kb encompasses genes from CmoCh08G001670 to CmoCh08G001830 of *C. moschata*. Comparing the orientations of this syntenic block, the physical positions of genes in both genomes are reversed. Inversions are believed to play an important role in speciation and local adaptation by reducing recombination and protecting genomic regions from introgression (Yang L. et al., 2014). The cluster of genes within this syntenic region contains transcription factors that have been described to confer resistance to viruses in different crops.

Genes of the same family of the WRKY transcription factor-like protein of *C. moschata* (CmoCh08G001670) appears to be involved in defense responses upon TMV infection in *C. annuum* (Huh et al., 2012). In PI 604506, six 3′ UTR variants are affecting this gene. Moreover, a BZIP transcription factor gene (CmoCh08G001710) is placed close to SNP_8061105. Although CmoCh08G001470 is not placed in the syntenic region with *C. melo*, it also encodes a BZIP transcription factor gene. Particularly, a stop codon lost has been detected in this gene of PI 604506, which could alter the primary structure of the protein.

Two genes encoding MADS-box transcription factors are in this same region (CmoCh08G001750 and CmoCh08G001760). This gene family has been associated to different virus-resistance mechanisms. A MADS-box transcription factor was described as the *Ty-2* candidate, involved in the tomato resistance to *tomato yellow leaf curl virus* (TYLCV) (Yang X. et al., 2014), and recently, a MADS-box gene has been reported to be upregulated in the *Sw-7* resistance to *tomato spotted wilt tospovirus* (TSWV) (Padmanabhan et al., 2019). No SNPs with high-impact predicted effect were identified in CmoCh08G001760 between *C. moschata* accessions, but changes in 5′ and 3′ UTRs and a missense mutation with predicted moderate effect were polymorphic between resistant and susceptible accessions. This gene has no ortholog in *C. pepo* chromosome 17, likely due to the insertion affecting this region of the genome. The possible involvement of this gene in ToLCNDV resistance would explain the total susceptibility to ToLCNDV found within *C. pepo* species (Sáez et al., 2016) and the difficulties to introgress the resistance *locus* from *C. moschata* to *C. pepo*.

The CmoCh08G001760 gene has paralogs in different chromosomes of *C. moschata*. OrthoMCL detected eight putative paralogs in different chromosomes of *C. moschata* (Chr1, Chr8, Chr12, Chr14, Chr17, and Chr18). The alignment of the aa sequences of the *C. moschata* paralogs and all MADs-box genes of *Arabidopsis thaliana* shows that CmoCH08G001760.1 is most similar to the *C. moschata* paralog located in Chr17, CmoCh17G013780.1. Both genes clustered together and apart from *A. thaliana* genes (**Supplementary Figure 2**). The detailed comparison of the aa sequences of both genes showed significant length differences (169 aa versus 71 aa, for CmoCH08G001760 and CmoCh17G013780.1 respectively). Both proteins have a common MADs motif at the N-terminus of the protein but differ in the rest of the sequence. These results are consistent with a different function of both genes. The sequence comparison of the CmoCh17G013780.1 gene of both parentals, PI 604506 and PI 419083 (done using the genomic sequences available at NCBI under BioProject PRJNA604046), does not provide SNP variants between them, also supporting the absence of a role of this paralog in ToLNDV resistance.

Molecular markers located close to the QTLs detected here can be used in marker-assisted selection in breeding ToLCNDV-resistant pumpkins and squash. Further genetic and transcriptomic studies of the candidate genes for resistance to ToLCNDV in the different cucurbit sources of resistance analyzed to date, are needed to develop strategies to control virus useful in different species of this crop family.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CS, BP, and CL conceived and designed the research. CS, CL, and AS performed the tests with ToLCNDV. CS, CM, CE, and BP conducted the marker development and mapping analysis. JM-P, JB, and CS performed the bioinformatics analysis of the genomic variation and synteny. CS and BP conducted and wrote the manuscript with important contributions from JM-P, MF, and CL. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00207/full#supplementary-material

FIGURE S1 | Dot plot showing the alignment between the QTL region in the previous assembly and the new assembly. **(A)** Chromosome 8 of *C. moschata* assembly v.1 vs. scaffold 15 of the new assembly, **(B)** chromosome 17 of *C. pepo* v.4.1 and scaffold 17 of the new assembly and **(C)** syntenic region for *C. moschata* and *C. pepo* new assembly.

FIGURE S2 | (A) Maximum likelihood tree, built using IQ-TREE v.1.5.2 (Nguyen et al., 2015), of amino acid sequences of *Arabidopsis thaliana* MADs-box and CmoCh08G001760 paralogs (in bold). Bootstrap values higher than 0.5 and lower than 0.95 are shown in the tree. Nodes with bootstrap values lower than 0.5 have been collapsed. **(B)** Conserved motifs found in all the *C. moschata* paralogs of gene CmoCH08G001760 by MEME Suite v.5.0.3. Red box represent the MAD motif.

TABLE S1 | List of SNP markers polymorphic in the F2 population derived from the cross PI 419083 × PI 604506. Their positions in the *Cucurbita moschata* genome are according to the Version 1 (http://cucurbitgenomics.org/). The positions in the genetic map is according to the genetic map constructed with the F2 plants in this study and used for QTL analysis.

TABLE S2 | SNPs used in all the QTLs validations. The chromosome and genetic position in each *C. moschata* × *C. moschata* and *C. pepo* × *C. moschata* linkage maps is indicated. Genomic position is shown in the last version of the *C. moschata* (v1.0), C. pepo (v4.1) and C. melo (v3.6.1) genomes available in Cucurbit Genomics Database (http://cucurbitgenomics.org). Flanking sequence of all markers is provided as well as the oligonucleotides used in the HRM genotyping assay. Markers that failed to show the expected polymorphism between parental sequences are marked as no data (−). Query cover, identity and E-value of Blast alignments with chromosome 11 of *C. melo* is also provide. Markers without significant similarity with chromosome 11 of *C. melo*, considering a minimum overlap between sequences of 70%, are shown as not applicable data (n/a). Striped and dotted lines are enclosing the interval position of the QTLs identified in the F2 (PI 419083 × PI 604506) and F2 (MU-CU-16 × PI 604506), respectively.

TABLE S3 | List of SNPs with their predicted effect on gene function on candidate region for *C. moschata* parents and for the six transcriptomes of *C. moschata*. Gene affected, SNP location, reference and alternative alleles, allele causing the predicted high impact, type of change and genotype for each accession is shown. Note that genotypes are coded with number of allele, where 0 means the reference allele and 1,2,3 refer to the alternative allele in the displayed order. Genes with related function to plant virus resistance are marked with *.

TABLE S4 | Genes annotated in the candidate region of chromosome 8 of *C. moschata*. Their paralog genes in chromosome 17 of *C. moschata* are shown, as well as BLAST alignment results with chromosome 17 of *C. pepo*. Those genes with none paralogs copies identified along the *C. moschata* genome are marked with *.

TABLE S5 | Statistical significance of synteny between homologous genes in candidate region for ToLCNDV resistance in chromosome 8 of *C. moschata* and chromosome 11 of *C. melo* genomes v1.0 and v3.6.1, respectively. Genes shown in bold are indicating the pick LOD in both QTLs analysis Cucurbita populations and *C. melo*, respectively. Dotted lines are defining the candidate region of *C. melo* were the major locus, responsible of the resistance to ToLCNDV, was identified.

TABLE S6 | Annotation results and fasta sequence of the Long Terminal Repeats (LTR) retrotransposon in *C. pepo* genome. The region was annotated according to the methodology for annotate repetitive elements described at Montero-Pau et al. (2018).

REFERENCES

Bio-Rad Laboratories, (2006). *Real-Time PCR Applications Guide*. Hercules, CA: Bio-Rad Laboratories, Inc, 42–43.

Capuozzo, C., Formisano, G., Iovieno, P., Andolfo, G., Tomassoli, L., Barbella, M. M., et al. (2017). Inheritance analysis and identification of SNP markers associated with ZYMV resistance in *Cucurbita pepo*. *Mol. Breed.* 37:99. doi: 10.1007/s11032-017-0698-5

Chang, H. H., Ku, H. M., Tsai, W. S., Chien, R. C., and Jan, F. J. (2010). Identification and characterization of a mechanical transmissible begomovirus causing leaf curl on oriental melon. *Eur. J. Plant Pathol.* 127, 219–228. doi: 10.1007/s10658-010-9586-0

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222. doi: 10.1093/bioinformatics/btv710

Cingolani, P., Platts, A., Wang, le L, Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Decker-Walters, D., and Walters, T. (2000). "Squash," in *The Cambridge World History of Food*, eds K. Kiple, and K. Ornelas (Cambridge, MA: Cambridge University Press), 335–351.

Dhillon, N. P. S., Monforte, A. J., Pitrat, M., Pandey, S., Singh, P. K., Reitsma, K. R., et al. (2012). Melon landraces of India: contributions

and importance. *Plant Breed.* 35, 85–150. doi: 10.1002/9781118100509.ch3

Díaz, J. A., Nieto, C., Moriones, E., and Aranda, M. A. (2002). Spanish *Melon necrotic spot virus* isolate overcomes the resistance conferred by the recessive *nsv* gene of melon. *Plant Dis.* 86:694. doi: 10.1094/PDIS.2002.86.6.694C

Diaz-Pendón, J. A., Trunger, V., Nieto, C., García-Mas, J., Bendahmane, A., and Aranda, M. A. (2004). Advances in understanding recessive resistance to plant viruses. *Mol. Plant Pathol.* 5, 223–233. doi: 10.1111/j.1364-3703.2004.00223.x

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

EPPO (2017). *European and Mediterranean Plant Protection Organization.* Available at: https://www.eppo.int/QUARANTINE/Alert_List/viruses/ToLCNDV.htm (accessed on 28 August 2017).

EPPO (2019). *European and Mediterranean Plant Protection Organization.* Available at https://gd.eppo.int/taxon/TOLCND/distribution (accessed on 20 November 2019).

FAO (2017). *FAOSTAT.* Available at: http://www.fao.org/faostat/en/#data/QC (accessed 5 February 2020).

Gallois, J. L., Moury, B., and German-Retana, S. (2018). Role of the genetic background in resistance to plant viruses. *Int. J. Mol. Sci.* 19:2856. doi: 10.3390/ijms19102856

Garcia-Ruiz, H. (2018). Susceptibility genes to plant viruses. *Viruses* 10:484. doi: 10.3390/v10090484

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* [Preprint]. arXiv:1207.3907

Giner, A., Pascual, L., Bourgeois, M., Gyetvai, G., and Rios, P. (2017). A mutation in the melon vacuolar protein sorting 41 prevents systemic infection of *Cucumber mosaic virus*. *Sci. Rep.* 7:10471. doi: 10.1038/s41598-017-10783-3

Haider, M. S., Tahir, M., Latif, S., and Briddon, R. W. (2006). First report of tomato leaf curl New Delhi virus infecting *Eclipta prostrata* in Pakistan. *Plant Pathol.* 55:285. doi: 10.1111/j.1365-3059.2005.01278.x

Hashimoto, M., Neriya, Y., Yamaji, Y., and Namba, S. (2016). Recessive resistance to plant viruses: potential resistance genes beyond translation initiation factors. *Front. Microbiol.* 7:1695. doi: 10.3389/fmicb.2016.01695

Huang, T. S., Wei, T., Laliberté, J. F., and Wang, A. (2010). A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant Physiol.* 152, 255–266. doi: 10.1104/pp.109.147983

Huh, S. U., Choi, L. M., Lee, G. J., and Kim, Y. J. (2012). *Capsicum annuum* WRKY transcription factor d (CaWRKYd) regulates hypersensitive response and defense response upon tobacco mosaic virus infection. *Plant Sci.* 197, 50–58. doi: 10.1016/j.plantsci.2012.08.013

Hussain, M., Mansoor, S., Iram, S., Zafar, Y., and Briddon, R. W. (2004). First report of tomato leaf curl New Delhi virus affecting chilli pepper in Pakistan. *Plant Pathol.* 53:794. doi: 10.1111/j.1365-3059.2004.01073.x

Ishibashi, K., Miyashita, S., Katoh, E., and Ishikawa, M. (2012). Host membrane proteins involved in the replication of tobamovirus RNA. *Curr. Opin. Virol.* 2, 693–698. doi: 10.1016/j.coviro.2012.09.011

Islam, S., Munshi, A. D., Mandal, B., Behera, T. K., and Kumar, R. (2010). Genetics of resistance in *Luffa cylindrical* Roemagainst tomato leaf curl New Delhi virus. *Euphytica* 174, 83–89. doi: 10.1007/s10681-010-0138-7

Islam, S., Munshi, A. D., Verma, M., Arya, L., Mandal, B., Behera, T. K., et al. (2011). Screening of *Luffa cylindrical* Roem for resistance against tomato leaf curl New Delhi virus, inheritance of resistance, and identification of SRAP markers linked to the single dominant resistance gene. *J. Hortic. Sci. Biotechnol.* 86, 661–667. doi: 10.1080/14620316.2011.11512819

Ito, T., Sharma, P., Kittipakorn, K., and Ikegami, M. (2008). Complete nucleotide sequence of a new isolate of tomato leaf curl New Delhi virus infecting cucumber, bottle gourd and muskmelon in Thailand. *Arch. Virol.* 153, 611–613. doi: 10.1007/s00705-007-0029-y

Jamil, N., Rehman, A., Hamza, M., Hafeez, A., Ismail, H., Zubair, M., et al. (2017). First report of tomato leaf curl New Delhi virus, a bipartite begomovirus, infecting soybean (*Glycine max*). *Plant Dis.* 101, 845. doi: 10.1094/PDIS-09-16-1267-PDN

Janssen, D., Simon, A., Crespo, O., and Ruiz, L. (2017). Genetic population structure of *Bemisia tabaci* in Spain associated with Tomato leaf curl New Delhi virus. *Plant Protect. Sci.* 53, 25–31. doi: 10.17221/62/2016-PPS

Joehanes, R., and Nelson, J. C. (2008). QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics* 24, 2788–2789. doi: 10.1093/bioinformatics/btn523

Juárez, M., Rabadán, M. P., Díaz Martínez, L., Tayahi, M., Grande-Pérez, A., and Gómez, P. (2019). Natural hosts and genetic diversity of the emerging tomato leaf curl New Delhi virus in spain. *Front. Microbiol.* 2019:140. doi: 10.3389/fmicb.2019.00140

Juárez, M., Tovar, R., Fiallo-Olivé, E., Aranda, M. A., Gosálvez, B., Castillo, P., et al. (2014). First detection of tomato leaf curl New Delhi virus infecting Zucchini in Spain. *Plant Dis.* 98, 857–858. doi: 10.1094/pdis-10-13-1050-PDN

Jyothsna, P., Haq, Q. M. I., Singh, P., Sumiya, K. V., Praveen, S., Rawat, R., et al. (2013). Infection of tomato leaf curl New Delhi virus (ToLCNDV), a bipartite begomovirus with betasatellites, results in enhanced level of helper virus components and antagonistic interaction between DNA B and betasatellites. *Appl. Microbiol. Biotechnol.* 97, 5457–5471. doi: 10.1007/s00253-012-4685-9

Kang, B. C., Yeam, I., and Jahn, M. M. (2005). Genetics of plant virus resistance. *Annu. Rev. Phytopathol.* 43, 581–621. doi: 10.1146/annurev.phyto.43.011205.141140

Kheireddine, A., Sifres, A., Sáez, C., Picó, B., and López, C. (2019). First report of tomato leaf curl New Delhi virus infecting cucurbit plants in Algeria. *Plant Dis.* 103:3291. doi: 10.1094/PDIS-05-19-1118-PDN

Kielbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Res.* 21, 487–493. doi: 10.1101/gr.113985.110

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., and Daly, M. J. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174–181. doi: 10.1016/0888-7543(87)90010-3

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint]. arXiv:1303.3997

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

Li, L., Stoecker, T. C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Lincoln, S., Daly, M., and Lander, E. (1992). *Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report*, 3rd Edn. Cambridge, MA: Whitehead Institute.

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

López, C., Ferriol, M., and Picó, M. B. (2015). Mechanical transmission of tomato leaf curl New Delhi virus to cucurbit germplasm: selection of tolerance sources in *Cucumis melo*. *Euphytica* 204, 679–691. doi: 10.1007/s10681-015-1371-x

McCreight, J. D., Wiutemantel, W. M., Natwick, E. T., Sinclair, J. W., Crosby, K. M., and Gómez-Guillamón, M. L. (2017). Recessive resistance to CYSDV in melon TGR 1551. *Acta Hortic.* 1151, 101–108. doi: 10.17660/ActaHortic.2017.1151.17

Millard, S. P. (2013). *EnvStats: An R Package for Environmental Statistics*. New York, NY: Springer.

Mnari-Hattab, M., Zammouri, S., Belkadhi, M. S., Bellon Doña, D., Ben Nahia, E., Hajlaoui, M. R., et al. (2015). First report of tomato leaf curl New Delhi virus infecting cucurbits in Tunisia. *New Dis. Rep.* 31:21. doi: 10.5197/j2044-05882015031021

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., et al. (2018). De-novo assembly of zucchini genome reveals a whole genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnol. J.* 16, 1161–1171. doi: 10.1111/pbi.12860

Moriones, E., Praveen, S., and Chakraborty, S. (2017). Tomato leaf curl New Delhi virus: an emerging virus complex threatening vegetable and fiber crops. *Viruses* 9:264. doi: 10.3390/v9100264

Munger, H. M., and Provvidenti, R. (1987). Inheritance of resistance to zucchini yellow mosaic virus in *Cucurbita moschata*. *Cucurbit Genet. Coop. Rep* 10, 8–81.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nicaise, V. (2014). Crop immunity against viruses: outcomes and future challenges. *Front. Plant Sci.* 5:660. doi: 10.3389/fpls.2014.00660

Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., et al. (2006). An *eIF4E* allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *Plant J.* 48, 452–462. doi: 10.1111/j.1365-313X.2006.02885.x

Orfanidou, C. G., Malandraki, I., Beris, D., Kektsidou, O., Vassilakos, N., Varveri, C., et al. (2019). First report of tomato leaf curl New Delhi virus in zucchini crops in Greece. *J. Plant Pathol.* 101:799. doi: 10.1007/s42161-019-00265-y

Pachner, M., Paris, H. S., and Lelley, T. (2011). Genes for resistance to zucchini yellow mosaic in tropical pumpkin. *J. Hered.* 102, 330–335. doi: 10.1093/jhered/esr006

Pachner, M., Paris, H. S., Winkler, J., and Lelley, T. (2015). Phenotypic and marker-assisted pyramiding of genes for resistance to zucchini yellow mosaic virus in oilseed pumpkin (*Cucurbita pepo*). *Plant Breed.* 134, 121–128. doi: 10.1111/pbr.12219

Padidam, M., Beachy, R. N., and Fauquet, C. M. (1995). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *J Gen. Virol.* 76, 25–35. doi: 10.1099/0022-1317-76-1-25

Padmanabhan, C., Ma, Q., Shekasteband, R., Stewart, K. S., Hutton, S. F., Scott, J. W., et al. (2019). Comprehensive transcriptome analysis and functional characterization of PR-5 for its involvement in tomato Sw-7 resistance to tomato spotted wilt tospovirus. *Sci. Rep.* 9:7673. doi: 10.1038/s41598-019-44100-x

Panno, S., Iacono, G., Davino, M., Marchione, S., Zappardo, V., Bella, P., et al. (2016). First report of tomato leaf curl New Delhi virus affecting zucchini squash in an important horticultural area of southern Italy. *New Dis. Rep.* 33:6. doi: 10.5197/j2044-05882016033006

Paris, H. S., and Cohen, S. (2000). Oligogenic inheritance for resistance to zucchini yellow mosaic virus in *Cucurbita pepo*. *Ann. Appl. Biol.* 136, 209–214. doi: 10.1111/j.1744-7348.2000.tb00027.x

Paris, H. S., Cohen, S., Burger, Y., and Joseph, R. (1988). Single-gene resistance to zucchini yellow mosaic virus in *Cucurbita moschata*. *Euphytica* 37, 27–29. doi: 10.1007/BF00037219

Pratap, D., Kashikar, A. R., and Mukherjee, S. K. (2011). Molecular characterization and infectivity of a tomato leaf curl New Delhi virus variant associated with newly emerging yellow mosaic disease of eggplant in India. *Virol. J.* 8, 305. doi: 10.1186/1743-422X-8-305

Rai, N. K., Sahu, P. P., Gupta, S., Reddy, M. K., Ravishankar, K. V., Singh, M., et al. (2013). Identification and validation of an ISSR marker linked to Tomato leaf curl New Delhi virus resistant gene in a core set of tomato accessions. *Vegetable Sci.* 40, 1–6.

Romay, G., Pitrat, M., Lecoq, H., Wipf-Scheibel, C., Millot, P., Girardot, G., et al. (2019). Resistance against melon chlorotic mosaic virus and tomato leaf curl New Delhi virus in melon. *Plant Dis.* 103, 2913–2919. doi: 10.1094/PDIS-02-19-0298-RE

Rosen, R., Kanakala, S., Kliot, A., Pakkianathan, B. C., Farich, B. A., Santana-Magal, N., et al. (2015). Persistent, circulative transmission of begomoviruses by whitefly vectors. *Curr. Opin. Virol.* 15, 1–8. doi: 10.1016/j.coviro.2015.06.008

Roy, A., Spoorthi, P., Panwar, G., Bag, M. K., Prasad, T. V., Kumar, G., et al. (2013). Molecular evidence for occurrence of tomato leaf curl New Delhi virus in ash gourd (*Benincasa hispida*) germplasm showing a severe yellow stunt disease in India. *Indian J. Virol.* 24, 74–77. doi: 10.1007/s13337-012-0115-y

Sáez, C., Esteras, C., Martínez, C., Ferriol, M., Narinder, P. S. D., López, C., et al. (2017). Resistance to tomato Leaf Curl New Delhi Virus in melon is controlled by a major QTL located in chromosome 11. *Plant Cell Rep.* 36, 1571–1584. doi: 10.1007/s00299-017-2175-3

Sáez, C., Martínez, C., Ferriol, M., Manzano, S., Velasco, L., Jamilena, M., et al. (2016). Resistance to tomato leaf curl New Delhi virus in *Cucurbita* spp. *Ann. Appl. Biol.* 169, 91–105. doi: 10.1111/aab.12283

Sifres, A., Sáez, C., Ferriol, M., Selmani, E., Riado, J., Picó, B., et al. (2018). First Report of tomato leaf curl New Delhi virus infecting Zucchini in morocco. *Plant Dis.* 102:1045. doi: 10.1094/PDIS-10-17-1600-PDN

Singh, A. K., Mishra, K. K., Chattopadhyay, B., and Chakraborty, S. (2009). Biological and molecular characterization of a begomovirus associated with yellow mosaic vein mosaic disease of pumpkin from Northern India. *Virus Genes* 39, 359–370. doi: 10.1007/s11262-009-0396-4

Sohrab, S. S., Karim, S., Varma, A., Abuzenadah, A. M., Chaudhary, A. G., Damanhouri, G. A., et al. (2013). Characterization of tomato leaf curl New

Delhi virus infecting cucurbits: evidence for sap transmission in a host specific manner. *Afr. J. Biotechnol.* 12, 5000–5009. doi: 10.5897/AJB2013.12012

Sohrab, S. S., Mandal, B., Pant, R. P., and Varma, A. (2003). First report of association of tomato leaf curl virus-New Delhi with yellow mosaic disease of *Luffa cylindrica* in India. *Plant Dis.* 87:1148. doi: 10.1094/PDIS.2003.87.9.1148A

Srivastava, A., Kumar, S., Jaidi, M., Raj, S. K., and Shukla, S. K. (2016). First report of tomato leaf curl New Delhi virus on opium poppy (*Papaver somniferum*) in India. *Plant Dis.* 100:232. doi: 10.1094/PDIS-08-15-0883-PDN

Srivastava, K. M., Hallan, V., Raizada, R. K., Chandra, G., Singh, B. P., and Sane, P. V. (1995). Molecular cloning of Indian tomato leaf curl virus genome following a simple method of concentrating the supercoiled replicative form of viral DNA. *J. Virol. Methods* 51, 297–304. doi: 10.1016/0166-0934(94)00122-w

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid cucurbita genomes. *Mol. Plant.* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003

Sundararaj, D., Jesse, M., Gunasekaran, D., Riyaz, M., Thangavelu, R., Kathiravan, K., et al. (2019). First report of tomato leaf curl New Delhi virus infecting *Crossandra infundibuliformis* (L.) in India. *Plant Dis.* 104. doi: 10.1094/PDIS-08-19-1764-PDN

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596

Usharani, K. S., Surendranath, B., Paul-Khurana, S. M., Garg, I. D., and Malathi, V. G. (2004). Potato leaf curl – a new disease of potato in northern India caused by a strain of tomato leaf curl New Delhi virus. *Plant Pathol.* 53:235. doi: 10.1111/j.0032-0862.2004.00959.x

Van Ooijen, J. W. (2009). *MapQTL® 6 Software for the Mapping of Quantitative Trait Loci in Experimental Population of Diploid Species.* Wageningen: Kayzma BV.

Vossen, R. H., Aten, E., Roos, A., and den Dunnen, J. T. (2009). High-resolution melting analysis (HRMA) - More than just sequence variant screening. *Hum. Mutat.* 30, 860–866. doi: 10.1002/humu.21019

Wessel-Beaver, L. (2005). Cultivar and germplasm release. Release of 'Soler' tropical pumpkin. *J. Agric. Univ.* 89, 263–266.

Yang, L., Koo, D., Li, D., Zhang, T., Jiang, J., Luan, F., et al. (2014). Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in *Cucumis*. *Plant J.* 77, 16–30. doi: 10.1111/tpj.12355

Yang, X., Yang, X., Caro, M., Hutton, S., Scott, J., Guo, Y., et al. (2014). Fine mapping of the tomato yellow leaf curl virus resistance gene Ty-2 on chromosome 11 of tomato. *Mol. Breed.* 34, 749–760. doi: 10.1007/s11032-014-0072-9

Zaidi, S. S., Shafiq, M., Amin, I., Scheffler, B. E., Scheffler, J. A., Briddon, R. W., et al. (2016). Frequent occurrence of tomato leaf curl New Delhi virus in cotton leaf curl disease affected cotton in Pakistan. *PLoS One* 1:e0155520. doi: 10.1371/journal.pone.0155520

Zaidi, S. S., Shakir, S., Malik, H. J., Farooq, M., Amin, I., and Mansoor, S. (2017). First Report of tomato leaf curl New Delhi virus on *Calotropis procera*, a weed as potential reservoir begomovirus host in Pakistan. *Plant Dis.* 101, 1071–1072. doi: 10.1094/PDIS-10-16-1539-PDN

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468.

Check for updates

QTL Mapping and Marker Development for Tolerance to Sulfur Phytotoxicity in Melon (*Cucumis melo*)

Sandra E. Branham[1,2], James Daley[3], Amnon Levi[1], Richard Hassell[2] and W. Patrick Wechter[1*]

[1] U.S. Vegetable Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Charleston, SC, United States, [2] Coastal Research and Education Center, Clemson University, Charleston, SC, United States, [3] HM.CLAUSE, Davis, CA, United States

Elemental sulfur is an effective, inexpensive fungicide for many foliar pathogens, but severe phytotoxicity prohibits its use on many melon varieties. Sulfur phytotoxicity causes chlorosis and necrosis of leaf tissue, leading to plant death in the most sensitive lines, while other varieties have little to no damage. A high-density, genotyping-by-sequencing (GBS)-based genetic map of a recombinant inbred line (RIL) population segregating for sulfur tolerance was used for a quantitative trait loci (QTL) mapping study of sulfur phytotoxicity in melon. One major (qSulf-1) and two minor (qSulf-8 and qSulf-12) QTL were associated with sulfur tolerance in the population. The development of Kompetitive Allele-Specific PCR (KASP) markers developed across qSulf-1 decreased the QTL interval from 239 kb (cotyledons) and 157 kb (leaves) to 97 kb (both tissues). The markers were validated for linkage to sulfur tolerance in a set of melon cultivars. These KASP markers can be incorporated into melon breeding programs for introgression of sulfur tolerance into elite melon germplasm.

Keywords: melon, *Cucumis melo*, sulfur tolerance, quantitative trait loci mapping, Kompetitive Allele-Specific PCR, sulfur phytotoxicity, whole genome resequencing

INTRODUCTION

Elemental sulfur is widely used as an organic fungicide in fruit and vegetable crops for control of powdery mildew and rusts (Williams and Cooper, 2004). For cucurbits, sulfur is an inexpensive and effective method for controlling powdery mildew (*Podosphaera xanthii*) (Koller, 2010; Keinath and Dubose, 2012). Sulfur can be applied to plants by direct contact, diffusion through water, or as a vapor (Bent, 1967). The underlying fungicide mechanism of sulfur is not known, but the current hypothesis is that it permeates into the fungus and interferes with mitochondrial respiration (Cooper and Williams, 2004), resulting in the inhibition of conidial germination (Gogoi et al., 2013). The Fungicide Resistance Action Committee defines sulfur's mode of action as multi-site contact activity and is considered a low risk for pathogen resistance development.

Although sulfur is used on many cucurbits, including melon, phytotoxic reaction to sulfur can range from extremely sensitive resulting in death of the plant, to completely tolerant (Johnson and Mayberry, 1980; Perchepied et al., 2004; Gogoi et al., 2013). Sulfur phytotoxicity is manifested as necrosis and pronounced "burning" on the leaf tissue starting four days after dusting fruiting melon plants in field conditions (Johnson and Mayberry, 1980). In greenhouse conditions, vaporized sulfur causes symptoms in as little as 24 h post-application in highly susceptible melon lines.

The limited research on sulfur phytotoxicity in melon has focused on sulfur dust application for tolerance screening and QTL discovery (Johnson and Mayberry, 1980; Perchepied et al., 2004). A sulfur tolerance screen of 31 melon cultivars by Johnson and Mayberry (1980) described 23 cultivars as tolerant and 8 as susceptible. In another study, 236 melon accessions from around the world were screened for response to sulfur, with 47% exhibiting complete tolerance (Perchepied et al., 2004). Perchepied et al. (2004) successfully mapped one major and two minor QTL associated with sulfur tolerance in two recombinant inbred line (RILs) populations sharing a common tolerant parent. The sulfur tolerance allele (contributed by the tolerant parent 'Vedrantais') of the major QTL exerted complete dominance when crossed to PI124112 and incomplete dominance to PI161375. The two minor QTL were only detected in the Vedrantais × PI124112 population. Perchepied et al. (2004) used a previously published genetic map (Périn et al., 2002) that was limited by low marker density (460 markers) resulting in poor QTL resolution, with the major QTL spanning 21 cM. The physical position of the QTL was not reported as there was not yet a melon reference genome available.

In this study, we utilized the high-density, genome-anchored genetic map available for the MR-1 × AY RIL population (Branham et al., 2018) to identify QTL associated with tolerance to vaporized sulfur. In addition, PCR-based markers for the major QTL were developed and tested in the population and various elite germplasm for linkage to sulfur tolerance and should be useful to breeders utilizing a marker assisted breeding scheme to increase the efficiency of sulfur tolerance introgression into elite cultivars.

METHODS

Experimental Design

A previously described RIL population (Branham et al., 2018) consisting of 170 lines generated from a cross of MR-1 and Ananas Yok'neum (AY) was evaluated for elemental sulfur tolerance. The Israeli cantaloupe cultivar Ananas Yok'neum was the sulfur tolerant parent and the inbred *C. melo* line MR-1 (Thomas, 1986) was the sensitive parent (**Figure 1**). Two independent greenhouse tests of the parents and population were initiated in May and June 2017. Each test was planted in a randomized complete block design with two replicates of five plants each. Lines were seeded into Metromix 360 (Sun Gro Horticulture, Agawam, MA) in 50-cell propagation trays (Hummert International, Earth City, MO) and allowed to grow to the 2–3 fully expanded leaf stage in a sulfur-free glass greenhouse. Temperature of the greenhouse were maintained at 30°C ± 5°C. Seedlings were fertilized the day prior to sulfur treatment by soaking trays in a liquid fertilizer solution (3 g Peters water soluble fertilizer per liter) (Scotts, Maryville, OH, USA). Trays were transferred into a temperature-controlled, 650 m^3 glass greenhouse for sulfur treatment. Temperature of the greenhouse were maintained at 30°C ± 5°C. Elemental sulfur (Soil Sulfur:>99% purity, National Garden Wholesale, Vancouver, WA, USA) was vaporized using two sulfur burners (Wilmod Sulfur Evaporator WSE75; Zoetermeer, Netherlands) for 4 hours nightly. The sulfur burners were ~2 m above the work benches, suspended 0.75 m below a circulation fan. The two burners were on adjacent ends of the greenhouse. Each sulfur burner vaporized approximately 1.2 g of sulfur per night. On the fifth day, lines were evaluated for sulfur tolerance by recording percent necrosis for both the most damaged cotyledon and true leaf on every plant (**Figure 2**). The percent necrosis for each RIL (cotyledon and true leaf) was averaged from evaluations of twenty plants (two tests × two reps × five plants). F$_1$ seeds failed to germinate in the original study, so an additional test was performed that included the parents and new seed of the F$_1$ hybrid. Two replicates of ten seeds per line were planted in a greenhouse trial in June 2018. An additional test of thirty melon accessions (cultivars and PIs) was evaluated to test the
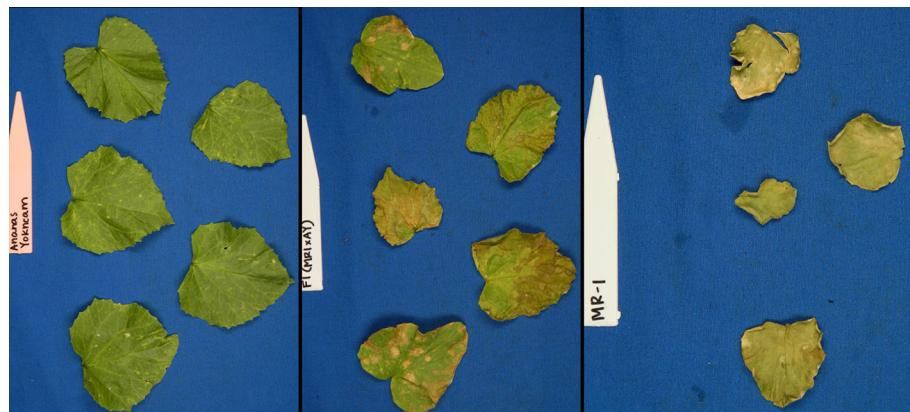
FIGURE 1 | Photographs of young plants of the sulfur tolerant (AY) and sensitive (MR-1) parents after vaporized elemental sulfur treatment.
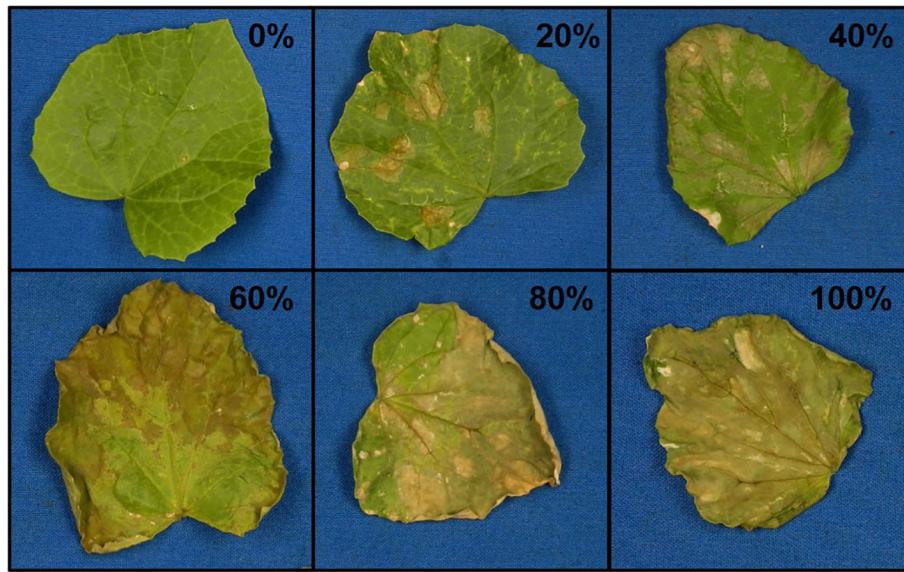
FIGURE 2 | Representative photos of damage ratings (0–100% in 20% increments) of melon leaves after vaporized elemental sulfur treatment.

utility of the sulfur markers in a variety of germplasm. Two replicates of five seeds each were planted in a greenhouse trial in March 2019. These additional studies followed the same protocols described above.

Statistical Analysis

Pearson's correlation (r) of line means between tests and between tissue types was calculated with the stats package of R version 3.4.1 (R Core Team, 2018). Broad-sense heritability (H^2) of sulfur tolerance, measured as percent affected leaf area (chlorosis and/or necrosis), was determined separately for each tissue type as the RIL variance divided by the total variance in percentage affected tissue area using variance components estimated with a linear mixed model in ASReml-R v3.0 (Gilmour et al., 2009). The model included RIL, test, interaction of RIL and test, replicate, and tray nested within test as random effects.

QTL Mapping

We used the previously published (Branham et al., 2018) high-density genetic map developed for this population for all QTL mapping analyses, which included 5,663 imputed, binned SNPs across the 12 chromosomes (=linkage groups) of the *C. melo* genome (Garcia-Mas et al., 2012). Haley–Knott regression (Haley and Knott, 1992) was used for multiple QTL mapping (MQM) with the stepwiseqtl function (Zeng et al., 1999; Broman and Speed, 2002; Broman and Sen, 2009) of Rqtl (Broman et al., 2003). The optimal QTL model based upon penalized LOD score (Manichaikul et al., 2009) was chosen through an automated forward and backward search algorithm. One thousand permutations of a two-dimensional, two QTL scan were used to calculate penalties and the genome-wide significance threshold for QTL detection. Multiple QTL models were visualized through LOD profile plots generated from forward

selection using standard interval mapping with Haley–Knott regression (Haley and Knott, 1992). Distributions of necrosis percentage of both cotyledons and true leaves did not meet the assumptions of parametric interval mapping, therefore the non-parametric model of the scanone function (Kruskal and Wallis, 1952; Kruglyak and Lander, 1995) was used for QTL verification. Genes within the 1.5-LOD interval of the major QTL were identified using the functional annotation of the *C. melo* reference genome v3.5.1 (Garcia-Mas et al., 2012), which was obtained through batch query at http://cucurbitgenomics.org/ (Zheng et al., 2019). In addition to using the functional annotation provided with the reference genome to search for candidate genes, conserved domains of genes were identified using the National Center for Biotechnology Information's batch CD search (CDDv3.16 database) (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2011; Marchler-Bauer et al., 2015; Marchler-Bauer et al., 2017).

Parental Resequencing

Genomic DNA was extracted from young leaf tissue of both parental lines (MR-1 and AY) using a DNeasy Plant Mini kit (Qiagen, Venlo, Netherlands) and sent to the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign for whole-genome resequencing. A Hyper Library construction kit (Kapa Biosystems, Roche, Basel, Switzerland) was used to prepare shotgun libraries for each parental DNA. Libraries were quantified by qPCR, pooled, and sequenced on one lane of a NovaSeq 6000 (Illumina, San Diego, CA) with a NovaSeq S2 reagent kit. Paired-end reads (150 bp) were demultiplexed with bcl2fastq v2.20 Conversion software (Illumina). Adaptors were trimmed from the 3' end of the reads. Duplicated read pairs were removed with perl scripts (https://github.com/Sunhh/NGS_data_processing/blob/master/drop_dup_both_end.pl). Low quality reads were removed with trimmomatic v0.38 (Bolger et al., 2014).

The remaining high-quality reads were aligned to *C. melo* reference genome v3.5.1 (Garcia-Mas et al., 2012) with BWA v0.7.17 (Li and Durbin, 2009). Picard v2.18.7 (http://broadinstitute.github.io/picard) was used to assign reads to a read group, tag reads originating from a single DNA fragment, and to create a reference sequence dictionary. The reference genome was indexed with Samtools v0.1.8 (Li et al., 2009). The Genome Analysis Toolkit (GATK v3.6) was used for SNP calling following the best practices for variant discovery (McKenna et al., 2010; Depristo et al., 2011, Van der Auwera et al., 2013). SNPs were filtered with Vcftools v0.1.15 (Danecek et al., 2011) to remove those with any missing data, heterozygous genotypes for either inbred parent, and/or genotype quality score of less than 30. SNPs within the major QTL region were functionally annotated with ANNOVAR version 2017 Jul 16 (Wang et al., 2010). Genes with missense or nonsense mutations and mutations to the promotor (less than 1 kb upstream of the start codon) were considered candidate genes.

Marker Development

The parental whole-genome resequencing data was used to design markers to saturate the region of the major sulfur tolerance QTL. Eighteen SNPs from across the major QTL region were developed into KASP markers (**Supplementary Table S1**) using "KASP^{TM} by design" services from LGC Genomics (Teddington, Middlesex, UK). PCR reactions (5 µl volume) consisted of 0.07 µl of primer mix (LGC Genomics; fluorophore-labeled allele-specific forward primers and a reverse primer), 2.5 µl of 2× master mix (LGC Genomics) and 20 ng of sample DNA. A standard thermal cycler was used for a touchdown PCR reaction with a 94°C hot-start activation step for 15 min, then ten cycles of 94°C (20 s) and a starting annealing temperature of 61°C that dropped by 0.6°C each cycle. Twenty-six additional cycles of 94°C for 20 s and 55°C for 60 s followed the touchdown steps. Fluorescence was quantified with a Stratagene Mx3005P (Agilent Technologies, Santa Clara, CA) quantitative PCR system at 25°C. Fluorescence values were used to cluster samples into genotypes with MxPro v4.10 software associated with the qPCR machine. Marker linkage to sulfur tolerance in the RIL population was assessed through QTL mapping both alone (KASP markers only) and combined with the binned GBS SNPs following the same procedures as described above. Thirty accessions (cultivars and plant introductions) were evaluated for sulfur tolerance and genotyped with the KASP markers. Correlation between the markers and sulfur phenotype of the accessions was assessed through analyses of variance (ANOVA) with the aov function (Chambers et al., 1992) in R.

RESULTS

Elemental Sulfur Tolerance

The population distribution of response to sulfur vapor, although strongly skewed towards tolerance with population means of 23.3 and 19.1% for cotyledons and leaves, respectively (**Figure 3**), varied widely across the population from 5.3 to 100% damage for cotyledons and 1 to 99.5% for leaves (**Supplementary Table S2**). The parents of the population had line means in the expected extremes of the distributions for both tissue types (**Figure 3**). AY
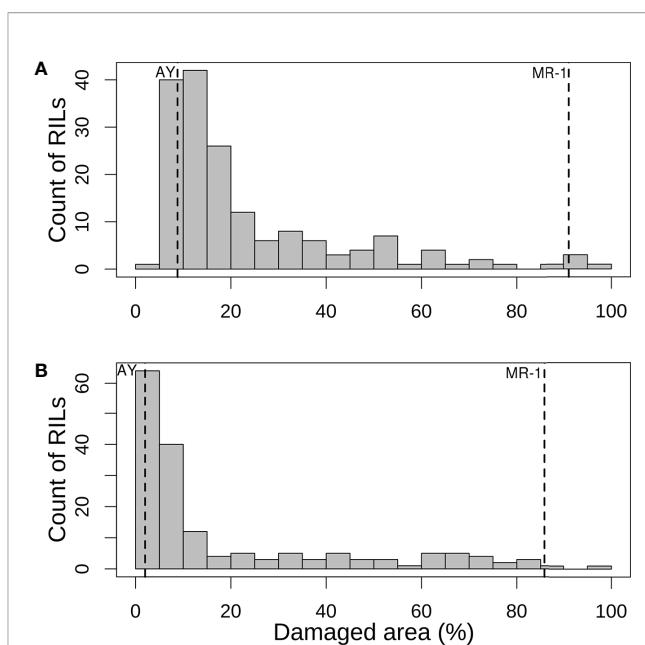
FIGURE 3 | Histograms of mean percentage of damaged (chlorotic or necrotic) area in **(A)** cotyledons and **(B)** leaves of the melon recombinant inbred line population after vaporized elemental sulfur treatment. Means of the tolerant parent (AY) and sensitive parent (MR-1) are indicated by vertical dashed lines.

had sulfur-induced damage of 8.8% for cotyledons and 2% for leaves, while MR-1 means were 91.5 and 85.4%, respectively. The RIL population means of affected area (chlorotic and/or necrotic) in response to sulfur treatment were highly correlated (p <2.2 × 10^{-16}; r = 0.85) between cotyledons and leaves. Correlation between tests was highly significant for both tissue types, but stronger for leaf tissue (p <2.2 × 10^{-16}; r = 0.83) than cotyledons (p <2.2 × 10^{-16}; r = 0.72). H^2 was moderate for both tissue types at 0.47 for cotyledons and 0.59 for leaves. An independent test of the parents and F$_1$ suggested dominant inheritance of sulfur sensitivity in the cotyledons (MR-1 = 97.8%, AY = 18.0% and F$_1$ = 93.0%), but incomplete dominance in the leaves (MR-1 = 90.3%, AY = 2.8% and F$_1$ = 75.1%). Mean percentage affected leaf area was higher in all tests and in all samples (parents, F$_1$, RILs) for cotyledons than for leaf tissue (**Table 1**).

QTL Mapping

A single major QTL (qSulf-1) on chromosome 1 explained 56.7 and 60.6% of the variation in mean sulfur tolerance across tests of cotyledons and leaves, respectively (**Figure 4**). RILs homozygous for the AY (tolerant) allele at qSulf-1 had a mean of 6.2% affected leaf area. A second minor QTL (qSulf-12) was associated with mean sulfur tolerance in cotyledons but not leaves (**Table 2**). The major QTL qSulf-1 is epistatic to qSulf-12. RILs homozygous for the AY (tolerant) allele at qSulf-1 show less than 15% damage, regardless of the genotype at qSulf-12 (**Supplementary Figure S1**). The sulfur tolerant allele for both QTL was contributed by the tolerant parent (AY).

MQM of cotyledon and leaf sulfur tolerance means for each test separately (tests 1 and 2) confirmed the location and major

Study	Experimental unit	Tissue	RIL population mean[a]	AY	MR-1	F_1[b]
RILs	Across	Cotyledon	23.4	8.8	91.1	NA
RILs	Test 1	Cotyledon	23.5	9.5	90.1	NA
RILs	Test 2	Cotyledon	23.4	8.0	92.2	NA
F_1	Across	Cotyledon	NA	18.0	97.8	93.1
RILs	Across	Leaf	19.3	2.0	85.8	NA
RILs	Test 1	Leaf	17.0	3.0	90.0	NA
RILs	Test 2	Leaf	21.7	1.0	81.1	NA
F_1	Across	Leaf	NA	2.8	90.3	75.3

[a,b]NA, Not applicable.

contribution (explained 42.1–65.7% of the variation in sulfur tolerance) of qSulf-1 (**Table 2**; **Supplementary Figure S2**). No minor QTL were associated with variation in leaf damage either across or within tests. The minor QTL identified for cotyledon damage, however, varied between the tests. Mean cotyledon damage across tests and in test 1 both identified qSulf-12 and the epistatic interaction with qSulf-1 (**Table 2**; **Supplementary Figure S2**). Sulfur tolerance means for test 2 were instead associated with a

new minor QTL, qSulf-8 (**Supplementary Figure S2**). The effect of this QTL was also masked by the AY allele from qSulf-1 but had a negative interaction (**Supplementary Figure S3**). The sulfur tolerant allele for qSulf-8 was contributed by the sulfur sensitive parent (MR-1). The Poisson distribution of sulfur response in the RIL population can be explained by the interaction of the QTL. RILs that were homozygous for the sulfur tolerant allele at qSulf-1 (N = 117 RILs) had mean damage of 14.0 and 6.3% in the cotyledons and leaves, respectively and represent the strong skew towards sulfur tolerance. The remaining tail of the distribution is comprised of individuals homozygous for MR-1 alleles at qSulf-1, with spread of the tail explained by genotypes at the minor QTL. RILs homozygous for sulfur sensitivity alleles at qSulf-1 but sulfur tolerance alleles at one or both of the minor QTL had mean damage of 38.3% for cotyledons and 43.8% for leaves. RILs at the extreme sensitive end of the sulfur response distribution were homozygous for sulfur sensitivity alleles at all three QTL and had mean damage of 68.6 and 67.3% for cotyledons and leaves, respectively.

The distribution of sulfur tolerance in the population did not fit the assumptions of the MQM methods so non-parametric interval mapping was used to confirm QTL. Non-parametric QTL mapping of sulfur tolerance means of cotyledons and leaf
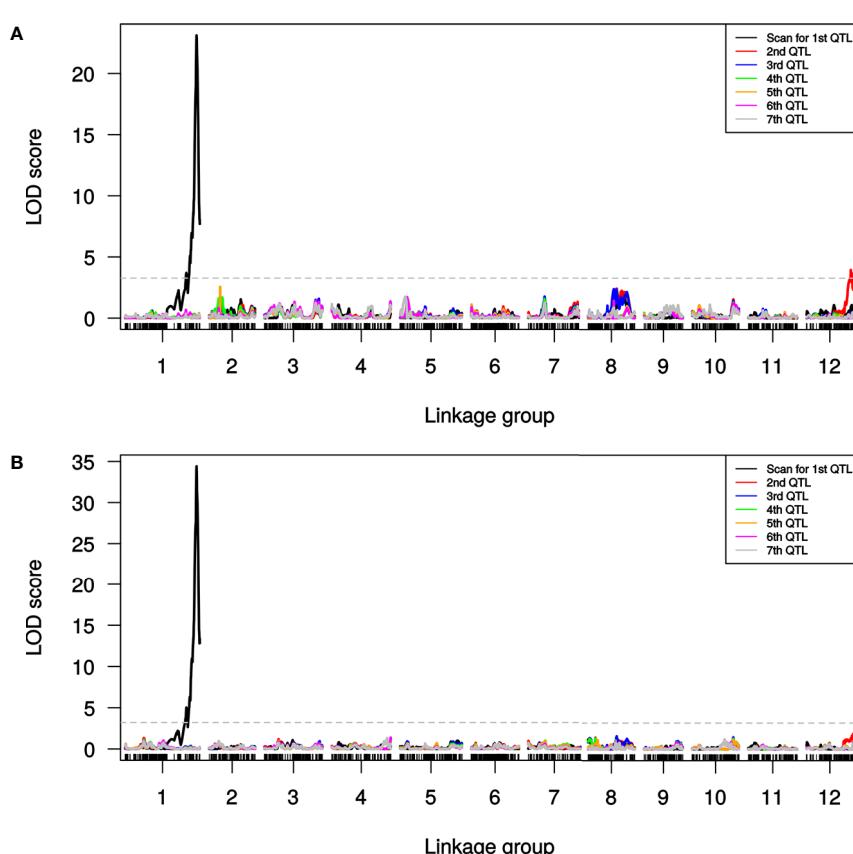
FIGURE 4 | Logarithm of odds (LOD) scores for forward model selection of up to seven QTL associated with mean percentage of damaged (chlorotic or necrotic) area of: **(A)** cotyledons and **(B)** leaves across two replicated greenhouse tests after vaporized elemental sulfur treatment. The initial scan shows the likelihood of the first QTL being located at each SNP in the genome (linkage group = chromosome) with subsequent scans showing the LOD of an additional QTL with the effects of the previous QTL(s) controlled for in the model. The dashed line marks the genome-wide significance threshold.

TABLE 2 | Quantitative trait loci (QTL) associated with sulfur tolerance means (percent affected tissue; cotyledon and leaf) within and across tests in a melon recombinant inbred line (RIL) population using different marker sets.

QTL[a]	Dataset[b]	Tissue	Peak(cM)	1.5-LODInterval (cM)	Total Interval (kb)[c]	LOD	%V$_P$[d]	P value	Additive[e]
qSulf-1	Across-KASP	Cotyledon	190.7	190.1–191	97	37.8	62.9	2.0×10^{-16}	−16.6
	Across		190.5	189.5–191.5	239	32.1	56.7	2.0×10^{-16}	−15.7
	Test 1		190.5	189–191.8	239	22.1	43.1	2.0×10^{-16}	−13.3
	Test 2		190.5	189–191.5	239	31.6	56.7	2.0×10^{-16}	−17.2
	Across-KASP	Leaf	190.7	190.1–191	97	49.5	74.0	2.0×10^{-16}	−22.3
	Across		190.5	189.6–191.5	157	34.4	60.6	0	−20.3
	Test 1		190	189–191.5	239	20.2	42.1	0	−13.4
	Test 2		190.5	189.6–191.5	157	38.3	65.7	0	−27.1
qSulf-12	Across-KASP	Cotyledon	117.7	117–119	174	12.9	14.7	2.3×10^{-13}	−7.5
	Across		117.7	116.8–121.4	410	9.9	12.6	2.0×10^{-10}	−6.6
	Test 1		118	117–119.3	174	9.3	15.0	8.8×10^{-10}	−7.6
	Across-KASP	Leaf	118	101.5–123.7	2,200	7.2	5.6	9.3×10^{-8}	−5.4
qSulf-8	Test 2	Cotyledon	68.4	66–69.2	360	8.9	11.3	2.0×10^{-9}	6.5
q1xq12	Across-KASP	Cotyledon	NA	NA	NA	8.2	8.7	1.3×10^{-9}	6.5
	Across		NA	NA	NA	6.0	7.2	2.3×10^{-7}	5.9
	Test 1		NA	NA	NA	5.1	7.9	1.6×10^{-6}	6.3
	Across-KASP	Leaf	NA	NA	NA	5.1	3.8	2.0×10^{-6}	−7.1
q1xq8	Test 2	Cotyledon	NA	NA	NA	6.5	7.9	7.4×10^{-8}	5.2

[a]QTL is named according to: q(trait abbreviation) – (chromosome = linkage group).
Epistatic interaction between qSulf-1 and qSulf-12 = q1xq12.
Epistatic interaction between qSulf-1 and qSulf-8 = q1xq8.
[b]Dataset used for QTL analysis: Means across two tests or means within independent tests; All genotypic data included the GBS SNPs and those listed as -KASP also included KASP markers.
[c]Physical distance of the genome corresponding to the 1.5-LOD interval of the QTL.
[d]Percent of the phenotypic variation explained by the QTL.
[e]Additive effect of the QTL.

tissue both across and within tests verified the association of qSulf-1 in all instances but found no minor QTL.

Marker Development

The major QTL qSulf-1 was identified in an area of the map with low SNP density, and collocated with only two binned SNPs. The remaining 157 kb did not have any SNPs from the GBS data. The closest SNPs flanking the QTL peak were 0.9 and 1.3 cM away. Therefore, the first objective for KASP design was to saturate the region to increase resolution of qSulf-1. The second objective was to design markers at regular intervals with decreasing frequency as distance from the QTL peak increased that would be able to track the breakage of linkage drag for future marker-assisted backcross selection. SNPs were identified and chosen for design from whole-genome resequencing data of the parents of the population, MR-1 and AY. Paired-end libraries were sequenced, generating 52.3 million reads for the sulfur tolerant parent (AY) and 44.5 million reads for the sulfur sensitive parent (MR-1). An initial set of 3.3 million SNPs was called between the parental genomes and then filtered to 304,864 SNPs. While the qSulf-1 interval had 154 SNPs between the parents, 18 were chosen to fill gaps in the original genetic map. Genotypes of the RIL population from the KASP markers were used for QTL mapping both alone and in combination with the original GBS population genotypes.

MQM using the 18 KASP markers identified the same QTL peak for both tissue types, located at 33,860,724 bp on chromosome 1 (**Figure 5**). Seven KASP markers were located in the 1.5-LOD interval of the qSulf-1 with an average genetic distance between markers of 0.20 cM. The remaining markers were located outside of the QTL region with frequency decreasing with genetic distance from the peak of qSulf-1 (**Figure 5**).

MQM of sulfur tolerance using the combined genotypic dataset (GBS and KASP; N = 5,681 SNPs) improved saturation and resolution of qSulf-1 for both tissue types. The physical distance of the 1.5-LOD interval decreased by 141 kb for cotyledons and 59 kb for leaf tissue (**Table 2**). The narrowed interval corresponded to the exact same physical positions for both tissue types, which extended from 33,776,147 to 33,873,503 bp (97,356 bp) on chromosome 1. In addition, qSulf-12 surpassed the significance threshold for the leaf phenotype in the combined genotypic dataset while it had not with GBS SNPs alone (**Table 2**).

Five KASP markers were tightly linked to qSulf-1 in the RIL population, including the peak LOD score of 49.54 at 190.74 cM (Sulf1_33860724) and a haplotype block of 4 SNPs (Sulf1-33791317, Sulf1-33804906, Sulf1-33835488 and Sulf1-33851209) that map 0.01 cM away (190.73 cM) with a LOD score of 49.46. The mean percent of affected leaf area for RILs homozygous for the sulfur tolerance allele at each of these SNPs was 6%.

The KASP markers were used to genotype thirty melon accessions to determine whether these markers could be successfully utilized in a variety of breeding programs (**Supplementary Table S3**). Nine of the eighteen markers (bolded in **Supplementary Table S3**) were significantly associated with the sulfur response of the accessions. Two markers had the strongest association, Sulf1-33791317 (R2 = 0.66) and Sulf1-33835488 (R2 = 0.72), but each had one sensitive accession with homozygous
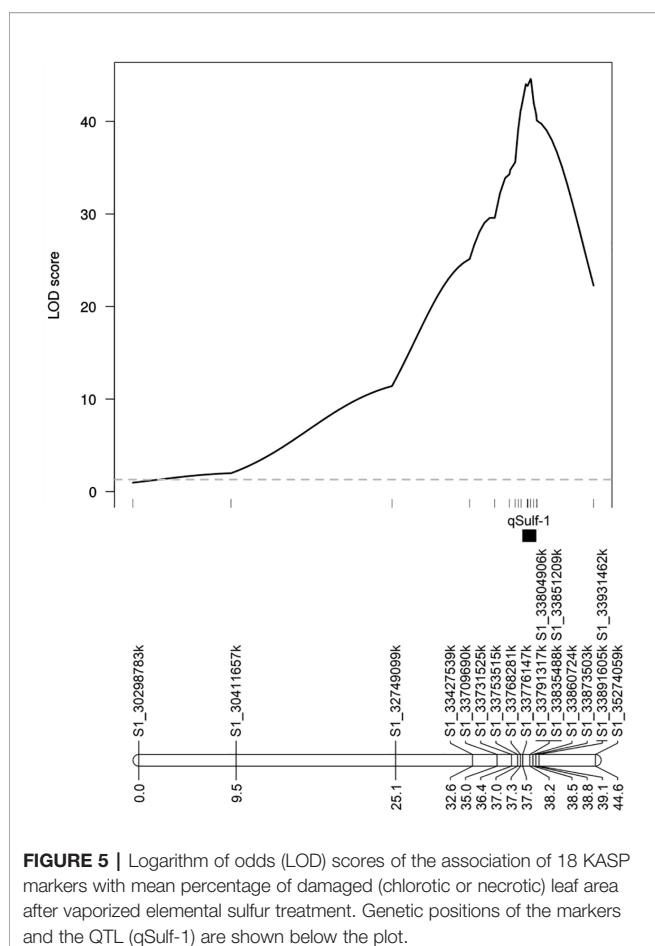
FIGURE 5 | Logarithm of odds (LOD) scores of the association of 18 KASP markers with mean percentage of damaged (chlorotic or necrotic) leaf area after vaporized elemental sulfur treatment. Genetic positions of the markers and the QTL (qSulf-1) are shown below the plot.

tolerant alleles. Both markers were also among the most tightly linked to sulfur response in the RIL population.

Candidate Genes

Candidate genes for sulfur tolerance in melon were identified by comparing the 1.5-LOD interval of qSulf-1 (GBS + KASP) with the physical location of the binned SNPs in the *C. melo* reference genome. Twenty-one genes (**Supplementary Table S4**) were encoded in the qSulf-1 interval (33,776,147 to 33,873,503 bp on chromosome 1). Whole genome sequencing identified 129 SNPs and 56 indels between the parents in the qSulf-1 region. Functional annotation of the polymorphisms narrowed the potential candidates to eight genes that had mutations most likely to cause regulatory or structural changes to the resulting proteins (**Supplementary Table S5**). Seven genes had SNPs or indels in the promoter region which may alter their expression levels. MELO3C024245 had two missense mutations in exon 5, altering the resulting amino acid sequence (**Supplementary Table S5**).

DISCUSSION

The strong skew towards tolerance in the response of the RIL population to vaporized elemental sulfur can be explained by the

epistatic interactions of the genes contributing to its polygenic inheritance. One major (qSulf-1) and two minor QTL (qSulf-8 and qSulf-12) were associated with sulfur tolerance in this study. RILs homozygous for the sulfur tolerance allele at the major QTL (qSulf-1) exhibited low damage independent of their genotype at the other loci. RILs homozygous for sulfur sensitivity alleles at qSulf-1 and tolerance alleles at one or both minor QTL displayed intermediate tolerance. RILs with sulfur sensitivity alleles at all three loci formed the extreme tail of the distribution.

Perchepied et al. (2004) found inheritance of sulfur tolerance to be polygenic in a QTL mapping study of sulfur tolerance in two sets of melon RILs with a common tolerant parent ('Vedrantais') and different sensitive parents (PI 161375 and PI 124112). Although the physical position of qSulf-1 cannot be compared to the QTL from this study, as the reference genome was not yet available, the chromosome names of the reference genome were based upon those of the linkage groups of the genetic map (Périn et al., 2002) used by Perchepied et al. (2004). Both studies found one major QTL on the proximal end of chromosome 1 with a similar contribution to variation in sulfur damage. Perchepied et al. (2004) also identified two minor QTL in one of the populations but they were found on different chromosomes than qSulf-8 and qSulf-12.

The limited research on tolerance to sulfur phytotoxicity in melon has focused on sulfur dust application for screening and QTL discovery (Johnson and Mayberry, 1980; Perchepied et al., 2004). Extensive research on the susceptibility of cucurbits to oxidized and reduced forms of sulfur may provide indications of the underlying mechanism of sulfur tolerance in some melon lines. Plants take in sulfur through their roots as sulfate and through their leaves primarily as sulfur dioxide and hydrogen sulfide, but excess sulfur accumulation can become toxic at levels that vary by species, varieties, soil-sulfur content, and environmental conditions (Rennenberg, 1984; Hawkesford and De Kok, 2006). Hydrogen sulfide can be oxidized by the plant to sulfate and reintroduced into the sulfur reduction pathway (Rennenberg and Filner, 1982), which through a series of enzymatic reactions produces cysteine, methionine, and glutathione (Hawkesford and De Kok, 2006). RILs in the melon population described in this report that are tolerant to sulfur toxicity may utilize the same pathways to expel excess elemental sulfur. One of the potential candidate genes based upon function includes a receptor-like protein kinase (RLK; MELO3C024237) that had eight polymorphisms in the promoter region. RLKs have been shown to play a critical role in plant responses to abiotic stresses in many studies (reviewed in Ye et al., 2017). A potential mechanism of sulfur tolerance may be that the RLK activates an enzyme in the sulfur metabolism pathway but lower expression of the RLK in MR-1 limits catabolism and sulfur accumulates. Ferredoxins provide the elections for sulfite reductase to reduce sulfite to sulfide, the precursor substrate for all products of the pathway (Hell, 1997). A gene encoding a ferredoxin (*MELO3C024264*) collocated with the original (GBS alone) qSulf-1 interval but was just outside the significant boundary after the addition of KASP markers. The sulfur sensitivity contributed by the MR-1 allele could be caused by lower expression of ferredoxin (due to an indel in the promoter region) limiting sulfur metabolism resulting in a buildup of sulfur to toxic levels.

MR-1 is known to be highly resistant to powdery mildew (Thomas, 1986), downy mildew (Thomas, 1986), Fusarium wilt (Branham et al., 2018), and *Alternaria* leaf blight (Thomas et al., 1990; Daley et al., 2017), making it an excellent source for resistance breeding. However, MR-1 is highly susceptible to both powdered and volatilized sulfur, thus introduction of sulfur susceptibility into a normally sulfur resistant elite line is a real concern. Here, we provide KASP markers tightly linked to the major sulfur tolerance QTL (qSulf-1), which can immediately be incorporated into melon breeding programs. These KASP markers will allow breeders to incorporate the disease resistances of MR-1 without the inadvertent introgression of sulfur susceptibility. MR-1 has poor horticultural quality in all traits (brix, texture, cracking, shape, etc.), therefore most breeding programs are likely to incorporate its many disease resistance alleles through backcrossing to an elite cultivar. Five of the markers released here are tightly linked to sulfur sensitivity in MR-1, while the remaining thirteen flank qSulf-1 with decreasing frequency. These markers will allow efficient tracking of introgression to limit linkage drag while ensuring exclusion of the major sulfur sensitivity allele of MR-1. In addition, two of the markers were significantly associated with sulfur response in a diverse set of cultivars suggesting they could be used to both avoid inadvertent introgression of sulfur sensitivity and introduce sulfur tolerance dependent upon the breeding materials chosen.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Dryad Data Repository, doi: 10.5061/dryad.zkh18937m.

ETHICS STATEMENT

The experiment conducted complies with the laws of the United States.

AUTHOR CONTRIBUTIONS

WW designed and implemented the experiments. JD optimized the sulfur protocols. WW and SB phenotyped the population. SB analyzed the data. SB, WW, AL, RH, and JD wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01097/full#supplementary-material

SUPPLEMENTARY TABLE S1 | Sequence information for the KASP markers, including: SNP ID, physical position of the SNP, primer sequences, SNP flanking sequence, and nucleotides of the sulfur tolerant (T) and sensitive (S) alleles.

SUPPLEMENTARY TABLE S2 | Phenotypic data used for QTL mapping: RIL means of percentage of damaged (chlorotic or necrotic) area after vaporized elemental sulfur treatment of cotyledons (cot) or leaves (leaf) in test 1 (_t1), in test 2 (_t2), and across tests (cot or leaf).

SUPPLEMENTARY TABLE S3 | Melon accession genotypes at KASP markers developed across qSulf-1. Markers are named according to physical position (bp) on chromosome 1. Genotypes are color coded, with individuals homozygous for the sulfur tolerance allele (**B**) in blue, sensitivity allele (**A**) in yellow, heterozygous (**H**) in gray and missing (**NA**) in white. The significance and magnitude of correlation between sulfur response and genotype are listed for each marker

SUPPLEMENTARY TABLE S4 | iChromosomal location and functional nformation for genes that collocated with the major QTL for sulfur tolerance, including the position of the start and stop codons within the chromosome (cs), gene ontology (GO) code and term, and conserved domains and features found within the gene (NCBI).

SUPPLEMENTARY TABLE S5 | Functional annotation of candidate gene polymorphisms between the parents in the QTL interval of qSulf-1, including the chromosome, physical position (in bp), parental alleles, polymorphism location relative to the gene (ie. upstream, exonic, downstream, etc.), distance from the gene (in bp), type of exonic mutation (synonymous or nonsynonymous), and detail (which exon and the nucleotide and amino acid changes).

SUPPLEMENTARY FIGURE S1 | Interaction plot showing evidence for epistasis between qSulf-1 and qSulf-12. Alleles from the sulfur sensitive parent (MR-1) are 'AA' and from the sulfur tolerant parent (AY) are 'BB'. The circle represents the mean percent damage of individuals in the population with the labelled genotypes. The plus signs indicate the standard error.

SUPPLEMENTARY FIGURE S2 | Logarithm of odds (LOD) scores for forward model selection of up to seven QTL associated with mean percentage of damaged (chlorotic or necrotic) area after vaporized elemental sulfur treatment of: (**a**) cotyledons in test 1, (**b**) cotyledons in test 2, (**c**) leaves in test 1 and (**d**) leaves in test 2. The initial scan shows the likelihood of the first QTL being located at each SNP in the genome (linkage group=chromosome) with subsequent scans showing the LOD of an additional QTL with the effects of the previous QTL(s) controlled for in the model. The dashed line marks the genome-wide significance threshold.

SUPPLEMENTARY FIGURE S3 | Interaction plot showing evidence for epistasis between qSulf-1 and qSulf-8. Alleles from the sulfur sensitive parent (MR-1) are 'AA' and from the sulfur tolerant parent (AY) are 'BB'. The circle represents the mean percent damage of individuals in the population with the labelled genotypes. The plus signs indicate the standard error.

REFERENCES

Bent, K. (1967). Vapour action of fungicides against powdery mildews. *Ann. Appl. Biol.* 60 (2), 251–263. doi: 10.1111/j.1744-7348.1967.tb04478.x

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

Branham, S. E., Levi, A., Katawczik, M., Fei, Z., and Wechter, W. P. (2018). Construction of a genome-anchored, high-density genetic map for melon (Cucumis melo L.) and identification of Fusarium oxysporum f. sp. melonis race 1 resistance QTL. *Theor. Appl. Genet.* 131, 829–837. doi: 10.1007/s00122-017-3039-5

Broman, K. W., and Sen, S. (2009). *A guide to QTL mapping with R/qtl (Vol. 46)* (New York: Springer).

Broman, K. W., and Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. R. Stat. Soc.: Ser. B. (Statistical Methodol.)* 64 (4), 641–656. doi: 10.1111/1467-9868.00354

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Chambers, J. M., Freeny, A., and Heiberger, R. M. (1992). "Analysis of variance; designed experiments," Chapter 5 of *Statistical Models in S*, Eds. S. J. M. Chambers and T. J. Hastie (Pacific Grove, CA: Wadsworth & Brooks/Cole).

Cooper, R. M., and Williams, J. S. (2004). Elemental sulphur as an induced antifungal substance in plant defence. *J. Exp. Bot.* 55 (404), 1947–1953. doi: 10.1093/jxb/erh179

Daley, J., Branham, S., Levi, A., Hassell, R., and Wechter, P. (2017). Mapping Resistance to Alternaria cucumerina in Cucumis melo. *Phytopathology* 107, 427–432. doi: 10.1094/PHYTO-06-16-0246-R

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

Depristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Gen.* 43, 491–498. doi: 10.1038/ng.806.A

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V. M., et al. (2012). The genome of melon (Cucumis melo L.). *Proc. Natl. Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109

Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2009). *ASReml User Guide Release 3.0* (Hemel Hemp- stead, UK: VSN International Ltd).

Gogoi, R., Singh, P. K., Kumar, R., Nair, K. K., Alam, I., Srivastava, C., et al. (2013). Suitability of Nano-sulphur for Biorational Management of Powdery mildew of Okra (Abelmoschus esculentus Moench) caused by Erysiphe cichoracearum. *J. Plant Pathol. Microbiol.* 4 (4), 171–175. doi: 10.4172/2157-7471.1000171

Haley, C. S., and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Hered. (Edinb)* 69, 315–324. doi: 10.1038/hdy.1992.131

Hawkesford, M. J., and De Kok, L. J. (2006). Managing sulphur metabolism in plants. *Plant Cell Environ.* 29 (3), 382–395. doi: 10.1111/j.1365-3040.2005.01470.x

Hell, R. (1997). Molecular physiology of plant sulfur metabolism. *Planta* 202, 138–148. doi: 10.1007/s004250050112

Johnson, J. R. H., and Mayberry, K. (1980). The effect of dusting sulfur on muskmelons. *HortScience* 15 (5), 652–654.

Keinath, A. P., and Dubose, V. B. (2012). Controlling powdery mildew on cucurbit rootstock seedlings in the greenhouse with fungicides and biofungicides. *Crop Prot.* 42, 338–344. doi: 10.1016/j.cropro.2012.06.009

Koller, M. (2010). Potassium Bicarbonate as a Potential Sulphur Substitute in Protected Organic Cropping. *I. Int. Conf. Organ. Greenhouse Horticult.* 915, 157–163.

Kruglyak, L., and Lander, E. S. (1995). A nonparametric approach for mapping quantitative trait loci. *Genetics* 139, 1421–1428.

Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47 (260), 583–621. doi: 10.1080/01621459.1952.10483441

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

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

Manichaikul, A., Moon, J. Y., Sen, Ś., Yandell, B. S., and Broman, K. W. (2009). A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* 181, 1077–1086. doi: 10.1534/genetics.108.094565

Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* 32, 327–331. doi: 10.1093/nar/gkh454

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39, 225–229. doi: 10.1093/nar/gkq1189

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C., Lu, S., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

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, 254–260. doi: 10.1101/gr.107524.110.20

Périn, C., Hagen, L. S., De Conto, V., Katzir, N., Danin-Poleg, Y., Portnoy, V., et al. (2002). A reference map of Cucumis melo based on two recombinant inbred line populations. *Theor. Appl. Genet.* 104, 1017–1034. doi: 10.1007/s00122-002-0864-x

Perchepied, L., Périn, C., Giovinazzo, N., Besombes, D., Dogimont, C., Pitrat, M., et al. (2004). Susceptibility to sulfur dusting and inheritance in melon. Progress in cucurbit genetics and breeding research. *Proc. Cucurbitaceae* 2004, 353–357.

R Core Team (2018). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Rennenberg, H., and Filner, P. (1982). Stimulation of h2s emission from pumpkin leaves by inhibition of glutathione synthesis. *Plant Physiol.* 69 (4), 766–770. doi: 10.1104/pp.69.4.766

Rennenberg, H. (1984). The fate of excess sulfur in higher plants. *Annu. Rev. Plant Physiol.* 35 (1), 121–153. doi: 10.1146/annurev.pp.35.060184.001005

Thomas, C., McCreight, J., and Jourdain, E. (1990). Inheritance of Resistance to Alternaria-Cucumerina in Cucumis-Melo Line Mr-1. *Plant Dis.* 74 (11), 868–870. doi: 10.1094/PD-74-0868

Thomas, C. (1986). Downy and Powdery Mildew Resistant Muskmelon Breeding Line Mr-1. *HortScience* 21 (2), 329–329.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43 (1), 11–10. doi: 10.1002/0471250953.bi1110s43

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164–e164. doi: 10.1093/nar/gkq603

Williams, J., and Cooper, R. (2004). The oldest fungicide and newest phytoalexin– a reappraisal of the fungitoxicity of elemental sulphur. *Plant Pathol.* 53 (3), 263–279. doi: 10.1111/j.0032-0862.2004.01010.x

Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J., and Zhu, C. (2017). The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep.* 36 (2), 235–242. doi: 10.1007/s00299-016-2084-x

Zeng, Z. B., Kao, C. H., and Basten, C. J. (1999). Estimating the genetic architecture of quantitative traits. *Genet. Res.* 74, 279–289. doi: 10.1017/S0016672399004255

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership