Theoretical and Applied Genetics (2024) 137:225
https://doi.org/10.1007/s00122-024-04741-7

ORIGINAL ARTICLE

Mapping and transcriptomic profiling reveal that the *KNAT6* gene is involved in the dark green peel colour of mature pumpkin fruit (*Cucurbita maxima* L.)

ChaoJie Wang[1,2] · Wenqi Ding[1,2] · Fangyuan Chen[1,2] · Ke Zhang[1,2] · Yuetong Hou[1,2] · Guichao Wang[1,2] · Wenlong Xu[1,2] · Yunli Wang[1,2] · Shuping Qu[1,2]

Received: 4 July 2024 / Accepted: 6 September 2024

Abstract

Key message **We identified a 580 bp deletion of *CmaKNAT6* coding region influences peel colour of mature *Cucurbita maxima* fruit.**

Abstract Peel colour is an important agronomic characteristic affecting commodity quality in *Cucurbit* plants. Genetic mapping of fruit peel colour promotes molecular breeding and provides an important basis for understanding the regulatory mechanism in *Cucurbit* plants. In the present study, the *Cucurbita maxima* inbred line '9-6' which has a grey peel colour and 'U3-3-44' which has a dark green peel colour in the mature fruit stage, were used as plant materials. At 5–40 days after pollination (DAP), the contents of chlorophyll a, chlorophyll b, total chlorophyll and carotenoids in the 'U3-3-44' peels were significantly greater than those in the '9-6' peels. In the epicarp of the '9-6' mature fruit, the presence of nonpigmented cell layers and few chloroplasts in each cell in the pigmented layers were observed. Six generations derived by crossing '9-6' and 'U3-3-44' were constructed, and the dark green peel was found to be controlled by a single dominant locus, which was named *CmaMg* (*mature green peel*). Through bulked-segregant analysis sequencing (BSA-seq) and insertion-deletion (InDel) markers, *CmaMg* was mapped to a region of approximately 449.51 kb on chromosome 11 using 177 F_2 individuals. Additionally, 1703 F_2 plants were used for fine mapping to compress the candidate interval to a region of 32.34 kb. Five coding genes were in this region, and *CmaCh11G000900* was identified as a promising candidate gene according to the reported function, sequence alignment, and expression analyses. *CmaCh11G000900* (*CmaKNAT6*) encodes the homeobox protein knotted-1-like 6 and contains 4 conserved domains. *CmaKNAT6* of '9-6' had a 580 bp deletion, leading to premature transcriptional termination. The expression of *CmaKNAT6* tended to increase sharply during the early fruit development stage but decrease gradually during the late period of fruit development. Allelic diversity analysis of pumpkin germplasm resources indicated that the 580 bp deletion in the of *CmaKNAT6* coding region was associated with peel colour. Subcellular localization analysis indicated that CmaKNAT6 is a nuclear protein. Transcriptomic analysis of the inbred lines '9-6' and 'U3-3-44' indicated that genes involved in chlorophyll biosynthesis were more enriched in 'U3-3-44' than in '9-6'. Additionally, the expression of transcription factor genes that positively regulate chlorophyll synthesis and light signal transduction pathways was upregulated in 'U3-3-44'. These results lay a foundation for further studies on the genetic mechanism underlying peel colour and for optimizing peel colour-based breeding strategies for *C. maxima*.

Communicated by Amit Gur.

ChaoJie Wang and Wenqi Ding have contributed equally to this work and should be considered co-first authors.

✉ Yunli Wang
wangyunli@neau.edu.cn

✉ Shuping Qu
spqu@neau.edu.cn

[1] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Northeast Region), Ministry of Agriculture and Rural Affairs/Northeast, Agricultural University, Harbin 150030, China

[2] College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin 150030, China

Published online: 17 September 2024

Introduction

Cucurbita is an economically important genus belonging to the Cucurbitaceae family and includes five major species: *Cucurbita pepo*, *Cucurbita maxima*, *Cucurbita moschata*, *Cucurbita argyrosperma, and Cucurbita ficifolia* (Chomicki et al. 2020; Ding et al. 2021). Peel colour is an important quality parameter used to measure fruit maturity and directly affects consumers' choice of pumpkin. There are many varieties of pumpkin fruits with different peel colours and patterns. The pumpkin peel is colourful, and the different varieties can have dark green, yellow−red, orange−red, and green skin with scattered yellow−red spots and other colours. Some peels have stripes with different widths and colours (Ezin et al. 2022; Kiramana and Isutsa 2017). The peel colour changes during fruit development. At the young fruit stage, the peel colour of *C. maxima* is mainly light yellow and light green, while the peel is different shades of green, red, and yellow at the mature fruit stage. The genetic mechanism of this characteristic is more complex than that of other crops, and a better understanding of the molecular genetics and gene mapping of peel colour in *C. maxima* are lacking.

Peel colour is related to the content of chlorophylls, carotene content, and flavonoids (Tadmor et al. 2010). Chlorophyll is necessary for photosynthesis, which gives plant organs a dark or light green colour depending on the chlorophyll content. Carotenoids and anthocyanins are the important pigments responsible for the yellow or red colour of fruits (Luo et al. 2019). Genes and quantitative trait loci (QTLs) associated with chlorophyll development and content controlling peel colour have been shown to be related to peel colour in tomato (*Solanum lycopersicum* L.), pepper (*Capsicum annuum* L.), cucumber (*Cucumis sativus*), and watermelon (*Citrullus lanatus*). The *STAY-GREEN* (*SGR*) mutation is a missense mutation that results in decreased chlorophyll content during fruit ripening in tomato (Barry et al. 2008). The *cl* mutation results in an amino acid substitution at an invariant residue in a pepper homologue of *SGR*, resulting in a reduction in the ripening-associated decrease in chlorophyll content (Barry et al. 2008). Genes involved in chlorophyll biosynthesis affect peel colour. The light green peel gene *CsYcf54*, which harbours the causative SNP, leads to a light green colour in cucumber (Lun et al. 2016). The dark green peel of watermelon is controlled by *ClCG08G017810*, whose coding region has a nonsynonymous SNP mutation, which results in a light green peel (Li et al. 2019). Mutations of carotenoid biosynthesis genes alter the carotene composition, thereby resulting in different fruit colours. *CYC-B* produces an orange colour by regulating the amount of β-carotene content in tomato (Hwang et al. 2016). A total of 1014 bp was inserted into the upstream promoter region of *LCY-E*, resulting in excessive production of δ-carotene to form a dark orange colour in tomato (Yoo et al. 2017). A nonsynonymous mutation in *CmOr* inhibits the downstream metabolism of *β*-carotene, participates in the transformation of chloroplasts to plastids, and regulates the peel colour of melon (Chayut et al. 2017). The major flavonoids in tomato fruits are quercetin and kaempferol, which are responsible for yellow fruit colour (Barros et al. 2012). *SlANT1* and *SlAN1* upregulate genes involved in the anthocyanin pathway, resulting in purple peel colour in the purple tomato lines (Kang et al. 2018). *CmKFB* negatively regulates the accumulation of naringenin chalcone, leading to yellow peel colour in melon (Feder et al. 2015).

Recent research has also revealed that many transcription factors control fruit ripening, which affects the accumulation of carotenoids, chlorophyll, and flavonoids. For example, the transcription factor Golden2-like (GLK) affects peel colour by regulating the chlorophyll content in tomato peels (Nguyen et al. 2014). The single-base insertion of *R2R3-MYB* in the inbred line WI7200 of cucumber results in a change in peel colour from orange to cream by regulating anthocyanin biosynthesis (Li et al. 2013). *CsMYB36* is involved in the formation of a yellow–green pericarp and regulates the colour of cucumber pericarp by regulating chlorophyll metabolism (Hao et al. 2018). *SlMYB12* plays a critical role in regulating flavonoids and leads to the production of pink tomatoes (Ballester et al. 2009; Fernandez-Moreno et al. 2016). A frameshift mutation and a stop codon of *BhAPRR2* in wax gourd lead to amino acid deletion and the formation of a green peel (Ma et al. 2021). A nonsynonymous mutation in *PRR2* (C2571T) resulted in a white pericarp colour, poor granular structure of the chloroplast, and a small plastid protein (Jeong et al. 2020). The *APRR2* gene is also present in cucumber, watermelon, and melon and regulates peel colour by affecting pigment synthesis and metabolism (Liu et al. 2015, 2016; Oren et al. 2019; Zhao et al. 2019). *KNOX* belongs to the THREE-AMINO-ACID LOOP-EXTENSION (TALE) class of transcription factors, which affect pericarp colour by regulating chloroplast development (Cortleven and Schmülling 2015; Nadakuduti et al. 2014; Tanaka et al. 2011). *TKN2* and *TKN4*, which encode tomato KNOX, are required for chloroplast development and distribution by upregulating *SlGLK2* and *SlAPRR2*-like expression at different latitudinal gradients, leading to a green shoulder in fruits (Nadakuduti et al. 2014). Accordingly, the pathway of KNOX genes in cucurbit crops is still unclear.

Previous studies have shown that many colour traits in *Cucurbita* are controlled by a single gene (Liu et al. 2020). *CmRc* regulates the orange rind in *C. maxima* and is closely related to SSR markers, with a genetic distance of 5.9 cM (Ge et al. 2015). The *pc* gene locus related to peel colour

was isolated from *C. moschata*, and light green peel was found to be dominant to dark green peel (Zhong et al. 2017). Kiramana and Isutsa (2017) reported that dominant *Gr* results in green and colouration and that the Bicolour *B-1* and *B-2* genes confer a precocious yellow colour in *C. pepo*. Based on whole-genome resequencing of *C. pepo* germplasms, *APRR*-2-like genes *Cp4.1LG05g02060* and *Cp4.1LG05g02070* were shown to regulate the yellow peel trait (Xanthopoulou et al. 2019). Through genetic mapping, *Cp4.1LG10g11560* (*CpCHLH*) was identified as a candidate gene for Y (yellow fruit colour), and its mutation resulted in a reduction in the chlorophyll content and yellow phenotype of *C. pepo* (Niu et al. 2023). However, detailed research on the molecular regulatory mechanisms related to peel colour in pumpkin species is lacking.

C. maxima has various peel colours; therefore, *C. maxima* is a valuable material for research on the mechanism of fruit peel colour. In our study, we performed genetic analysis and fine mapping based on bulked-segregant analysis sequencing (BSA-seq) combined with insertion-deletion (InDel) markers in two inbred lines ('9-6' and 'U3-3-44') with different peel colours. A new gene was identified as a promising candidate gene for controlling dark green peel colour in *C. maxima.* The transcriptomic data suggested that this gene may regulate chlorophyll development and biosynthesis by upregulating the expression of transcription factors involved in light signalling. This study provides a theoretical basis for further elucidating the mechanism underlying the formation of peel colour in pumpkin and breeds of brightly coloured varieties in pumpkin breeding.

Materials and methods

Plant materials

The parental inbred line of *C. maxima* with a grey peel '9-6' and one with a dark green peel 'U3-3-44' were used in the

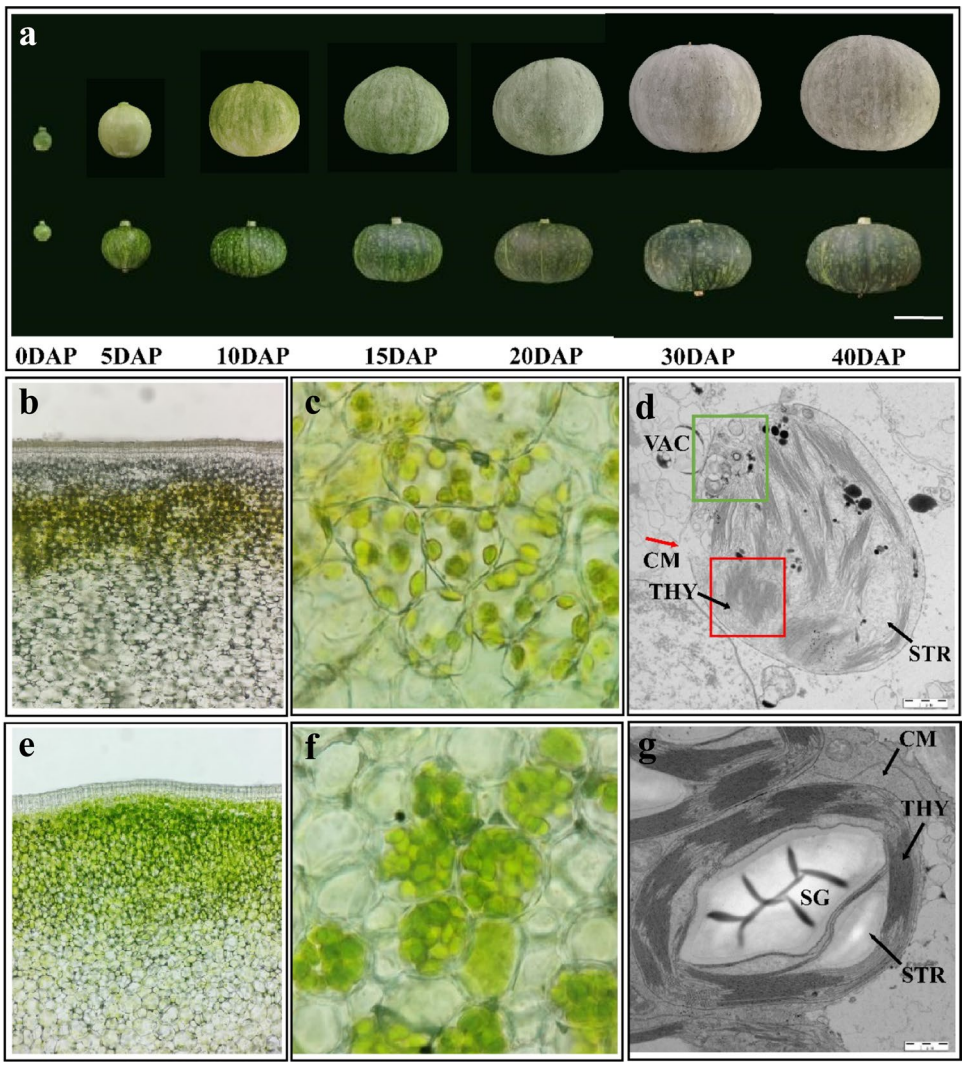

Fig. 1 Peel colour phenotype of '9-6' and 'U3-3-44' during fruit development. **a** Peel colour of '9-6' and 'U3-3-44' during 0–40 DAP. Bar = 10 cm. **b** Chloroplast distribution of epicarp in '9-6' at 30 DAP. **c** Chloroplast number of parenchymal cells in '9-6' epicarp at 30 DAP. **d** Chloroplast structures in '9-6' epicarp at 30 DAP. **e** Chloroplast distribution of epicarp in 'U3-3-44' at 30 DAP. **f** Chloroplast number of parenchymal cells in 'U3-3-44' epicarp at 30 DAP. **g** Chloroplast structures in 'U3-3-44' epicarp at 30 DAP. CM indicates chloroplast membrane, THY indicates thylakoid, STR indicates stroma, VAC indicates vacuole, and SG indicates starch grains

present study (Fig. 1). Both parents were globular in shape and flattened. The 9-6 × U3-3-44 hybridizations generated F_1 plants, which were self-pollinated to generate F_2 plants. Two backcrossed populations (BC_1P_1 and BC_1P_2) were generated by crossing the two parents with F_1 plants. The parental, F_1, F_2, BC_1P_1, and BC_1P_2 populations were used to assess and validate dark green/gray colour inheritance. F_2 fruits were used for linkage analysis and candidate gene identification. An additional 1703-F_2 segregating population and 517-BC_1 were used for mapping *CmaMg*. Then, $F_{2:3}$ lines derived from self-crossed F_2 plants and $F_{3:4}$ lines derived from self-crossed F_3 plants were generated for progeny testing to infer the *CmaMg* locus genotypes. For analysis of the allelic diversity of *CmaMg*, germplasm resources from 23 inbred lines of *C. maxima* were used to determine the associations between peel colour traits and *CmaMg* alleles. All plant materials were grown in a greenhouse at the Xiangyang Experimental Horticultural Station of Northeast Agricultural University (Harbin, China, N45°77′, E126°92′) from spring 2020 to 2022.

Phenotypic data collection

The parental, F_1, F_2, BC_1P_1, and BC_1P_2 plants had identification tags showing their serial numbers and pollination dates. The peel colour of the fruit was recorded at 6, 5, 4 days before pollination (DBP) and at 0–40 days after pollination (DAP), and each fruit was photographed for reassessment and determination of the chlorophyll content. The frequency distributions and significant differences of these data were identified using standard procedures in GraphPad Prism 8.0.2 and DPS 9.50 software.

Pigment content and chloroplast analysis

'9-6' and 'U3-3-44' as experimental materials, the epicarp was sampled at different developmental stages (0, 5, 10, 15, 20, 30, and 40 DAP), tissue samples of leaves and stems of the parental lines at 5 DAP were collected, and the pigments were extracted with 80% acetone for at least 24 h in darkness. A spectrophotometer (Epoch, Bio Tek Instruments Inc., USA) was used to measure the absorbance values of the chlorophyll a, chlorophyll b, and carotenoids at 663, 645, and 470 nm, respectively (Inskeep and Bloom 1985). Three replicates were prepared for each sample.

To compare the cytological characteristics and the chloroplasts of epidermal between the two parental lines, pericarp at 5–30 DAP was sliced with a sterile blade, placed on a glass slide with a small drop of distilled water, pressed onto a coverslip, made into microscopic sections (Lux et al. 2005), observed with light microscopy at 400 ×, and photographed. Pericarp at 30 DAP was fixed in 2.5% glutaraldehyde and observed by transmission electron microscope (H-7650; Hitachi, Tokyo, Japan) as described by Zhu et al. (2022).

Sequencing mapping strategy for BSA

Total genomic DNA was extracted from two parents and F_2 individual plants using the modified CTAB method (Porebski et al. 1997). Three young leaves were cut from the top of each individual and frozen in liquid nitrogen for subsequent DNA extraction. The parental dark green peel pool and parental grey peel pool were constructed by mixing an equal amounts of DNA from 10 '9-6' plants and 10 'U3-3-44' plants. Similarly, 60 individual plants (30 dark green and 30 grey) with extreme phenotypes were collected from the F_2 population to construct two DNA pools, the dark green-pool, and grey-pool. The qualified library was sequenced on the Illumina HiSeq 2500 platform. To ensure the quality of information analysis, we filtered the raw reads to obtain clean reads. Reads were mapped to the Cucurbit Genomics Database (*C. maxima*. cv. Rimu, http://cucurbitgenomics.org/ftp/genome/Cucurbita_maxima/) using BWA software (Li and Durbin 2009, https://bio-bwa.sourceforge.net). SNPs and InDels between the '9-6', 'U3-3-44', dark green-pool, and grey-pool were detected using GATK (McKenna et al. 2010, http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit). Two association analysis methods, the Euclidean distance (ED) algorithm and Δ (SNP /InDel index) algorithm, were used (Fekih et al. 2013; Hill et al. 2013). The candidate region with a trait association above the 99% threshold was identified.

Molecular marker development and genetic mapping

For preliminary mapping of *CmaMg*, InDel and KASP markers were developed based on InDels and SNPs in the two parental lines according to BSA-seq. The primers were designed with Primer Premier 5.0 and synthesized by the Beijing Genomics Institute (Table S1). InDel PCR amplification was performed in a 10 μL reaction with 1 μL DNA, 4 μL PCR master mix, 0.5 μL of 10 μM per primer, and 4 μL distilled water. The PCR program was as follows: initial denaturation at 94 °C for 5 min; 33 cycles of 94 °C for 30 s; 55 °C for 30 s, 72 °C for 30 s; 72 °C for 5 min; and holding at 4 °C. The amplicons were separated on an 8% polyacrylamide gel by phoresis. After electrophoresis at 220 V for 2 h, the gel was stained in 0.2% $AgNO_3$ solution and revealed the silver-stained DNA bands. KASP primers were designed according to the LGC Genomics database (http://www.lgcgenomics.com). PCRs were performed with a high-throughput hydrocycler system under the following conditions: 94 °C for 15 min; 10 cycles of 94 °C for 20 s, with a 0.6 °C reduction in each cycle until reaching 61 °C;

94 °C for 20 s; and 26 cycles of 55 °C for 60 s. After amplification was completed, the fluorescence signal was detected using a BMG PHERAstar instrument, and genotyping was performed with Kraken software. A total of 177 F_2 plants from '9-6' × 'U3-3-44' were initially used for the linkage analysis. Candidate intervals were determined by observing recombination events between the flanking markers. Genetic mapping was performed with JoinMap 4.0.

Sequence analysis and prediction of candidate genes

The sequence and gene function data were retrieved from the Cucurbit Genomics Database (Zheng et al. 2019) and the online program FGENESH (http://linux1.softberry.com/berry.phtml) (Salamov and Solovyev 2000). The function of the predicted genes was determined with the BLASTP tool from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The DNA sequence and the coding sequence (CDS) sequence of the candidate gene were cloned and inserted into the pEASY-T1 vector (TransGen, Beijing, China). The primer sequences were designed by Primer 5 software and are shown in Table S2.

RNA extraction and qRT-PCR analysis

Peel samples from different developmental stages (6, 5, 4 DBP and 0, 5, 10, 15, 20, 30, 40 DAP) and other tissue samples, including roots, stems, leaves, and fresh female and male flowers, were collected from both parental lines at 5 DAP. RNA was isolated by the TRIzol method (Invitrogen, USA). First-strand cDNA was synthesized using a cDNA synthesis kit (TOYOBO, Japan). Quantitative real-time PCR (qRT-PCR) was performed using TB Green® Premix Ex Taq™ II. The 20-µl qPCR volume included 2 µL of cDNA, 0.5 µL of each primer, 10 µL of TB Green Premix Ex Taq, and 7 µL of ddH_2O. The PCR protocol was as follows: 96 °C for 10 min; 40 cycles of 95 °C for 15 s, 60 °C for 60 s, and final extension at 72 °C for 30 s. Three replicates were used for each sample. Primers were designed by Primer Premier 5.0 software (Table S2). The *Actin* gene was used as the internal reference. Relative expression levels were determined with the $2^{-\Delta\Delta CT}$ method (Wang et al. 2022).

Allelic variation of *CmaMg* in germplasm resources

For validation of the relationship between the deletion sites of the candidate gene and peel colour, specific primers were designed according to the deletion site of the candidate gene through Primer Premier 5.0 (Table S2). The candidate gene was cloned and sequenced in '9-6', 'U3-3-44', and 23 inbred lines of *C. maxima*. The nucleotide sequences were aligned using DNAMAN 8.0 software, and germplasm resource information for *C. maxima* is listed in Table S3.

Phylogenetic analysis of *CmaMg* protein sequences

To investigate the phylogenetic relationships of the *CmaMg* protein in *C. maxima*, we selected all *KNAT* genes in *Arabidopsis thaliana* and *KNAT* homologues in *Oryza sativa, Zea mays, C. sativus, C. melo, C. lanatus, and S. lycopersicum*. The protein sequence of KNAT in *A. thaliana* was retrieved from TAIR (https://www.arabidopsis.org/), and the protein sequences of different species were obtained from NCBI (https://www.ncbi.nlm.nih.gov/) and Phytozome (https://phytozome-next.jgi.doe.gov/). A phylogenetic tree was generated using the neighbour-joining (NJ) method in MEGA 7.0 via bootstrap analysis with 1,000 replicate datasets, and the phylogenetic tree was constructed with itol software (https://itol.embl.de/login.cgi).

Subcellular localization of *CmaMg*

The full CDS of *CmaKNAT6* in the peel of 'U3-3-44' was cloned without the stop codon into a pCAMBIAsuper1300-GFP vector to generate a *35S*::CmaKNAT6-GFP fusion expression vector under the control of the cauliflower mosaic virus (CaMV) 35S promoter (Table S2). The recombinant vector was subsequently transformed into GV3101 and injected into tobacco leaves (Sparkes et al. 2006). The empty vector was used as a control. Green fluorescent protein (GFP) fluorescence was observed under a fluorescence microscope (Olympus BX53, Tokyo, Japan).

RNA extraction and RNA-Seq

Peel sample types from lines '9-6' and 'U3-3-44' were collected at 6 DBP and 20 DAP. The following sample types were labelled: G6 (gray peel at 6 DBP), G20 (gray peel at 20 DAP), DG6 (dark green peel at 6 DBP), and DG20 (dark green peel at 20 DAP). Three biological replicates were used for each sample type, and three plants were used for each replicate. Total RNA was extracted using TRIzol (TaKaRa, Otsu, Shiga, Japan) and treated with RNase-free DNase. mRNA was enriched by oligo (dT) magnetic beads, followed by random interruption of the mRNA in NEB fragmentation buffer (NEB #E6150S). cDNA was synthesized using random 6-base primers. After end repair, addition of an A-tail, and ligation of sequencing junctions, 250 bp of cDNA was screened with AMPure XP beads, amplified by PCR, and purified to obtain the cDNA library. RNA sequencing was performed by Novogene (Tianjin, China). A total of 12 libraries were sequenced on the Illumina platform (HiSeq 6000). Adapter sequences and low-quality bases were removed, and clean data were de novo assembled using the

Trinity platform (Sun et al. 2022). Clean reads were aligned to the reference genome of *C. maxima* 'Rimu' (http://cucurbitgenomics.org/ftp/genome/Cucurbita_maxima/v1.1/) using HISAT2 (Mortazavi et al. 2008), and de novo transcripts were assembled and quantified using StringTie software (Pertea et al. 2016, http://ccb.jhu.edu/software/stringtie). The read count numbers were calculated and translated to FPKM gene values (DeLuca et al. 2012). DEGs were analysed using DESeq2 (Anders and Huber 2010) and selected with $|\log2(\text{FoldChange})| \geq 1$ and padj ≤ 0.05. DEGs were mapped to gene ontology (GO) and Tokyo Encyclopedia of Genes and Genomes (KEGG) databases. Significantly enriched biochemical pathways were obtained with padj $\leq$ 0.05 after multiple testing to correct the P-value (Anders et al. 2010).

Results

Phenotype of the peel colour and pigment content

Ovaries and fruits of '9-6' and 'U3-3-44' at different developmental stages (0–6 DBP and 5–40 DAP) were collected. The ovary peel colours of '9-6' and 'U3-3-44' at 0–6 DBP were not significantly different (Fig. S1). The peel colours of '9-6' and 'U3-3-44' started to show obvious differences at 5 DAP (Fig. 1a). The fruit peel colour of '9-6' was light green, gradually became cream-coloured in the immature fruit stage (5–20 DAP), and finally turned grey in the mature fruit stage (30–40 DAP). During fruit development, the peel colour of 'U3-3-44' changed from light green to green (5–20 DAP) and finally gradually turned dark green (30–40 DAP). The colour of the organ on the ground of '9-6' and 'U3-3-44' was also observed, and the stem and leaves of '9-6' were lighter than those of 'U3-3-44'.

The chlorophyll and carotenoid contents in the '9-6' and 'U3-3-44' peels were determined (Fig. 2). The contents of chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids in 'U3-3-44' increased quickly in the immature fruit stage (0–30 DAP), while changed little in the mature fruit stage (30–40 DAP). In '9-6', the contents of chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids were low at 0–10 DAP and increased quickly at 10–30 DAP, after which the content increased gradually at 30–40 DAP. The contents of chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids in 'U3-3-44' were significantly greater than those in '9-6' during 5–40 DAP and approximately 4-, 5-, 4-, and 3-fold greater than those in '9-6' at 40 DAP, respectively. The changes in pigment content and the results of the naked-eye observations were consistent. We found that the stems and leaves colour of mature plants between 'U3-3-44' and '9-6' also differed (Fig S1). The stems and leaves colour

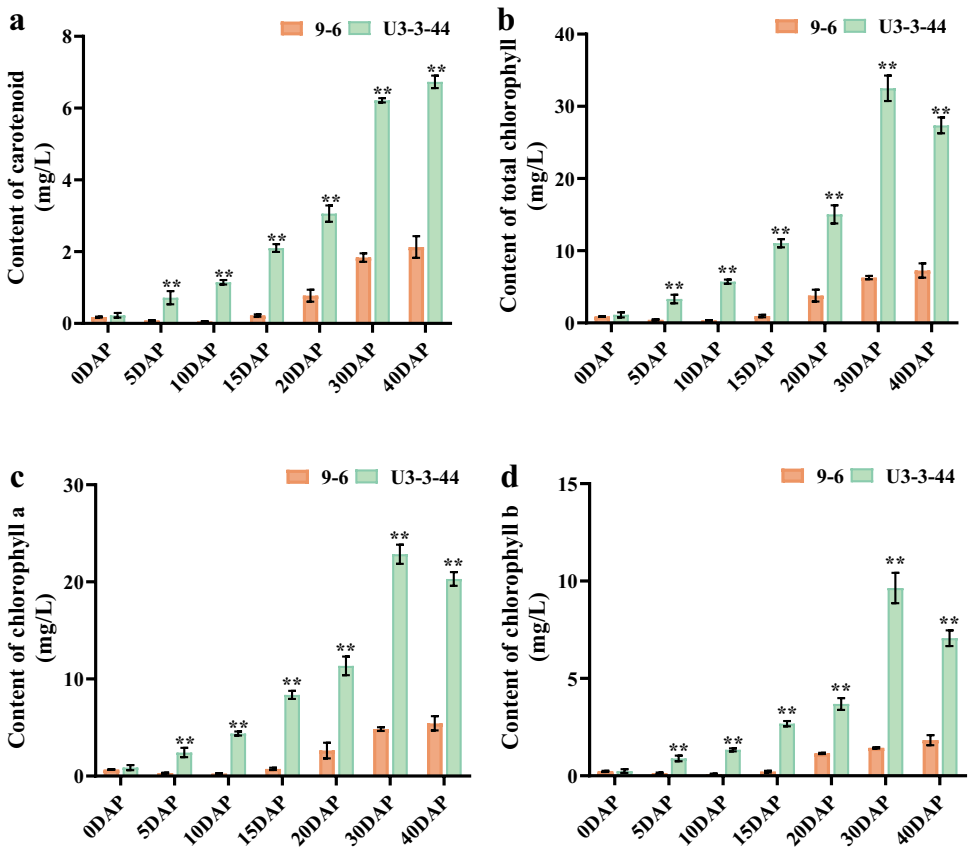

Fig. 2 Pigment content of '9-6' and 'U3-3-44' peel during fruit development. **a** Carotenoids content of '9-6' and 'U3-3-44' peel. **b** Total chlorophyll content of '9-6' and 'U3-3-44' peel. **c** Chlorophyll a content of '9-6' and 'U3-3-44' peel. **d** Chlorophyll b content of '9-6' and 'U3-3-44' peel. ** indicates an extremely significant difference, $P < 0.01$

in 'U3-3-44' were darker green than those in '9-6' (Fig S1b, d). The pigment contents of the stems and leaves were also measured (Fig S1e, f), and the results showed that the contents of chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids in 'U3-3-44' were significantly greater than those in '9-6'.

Observation of peel structures

Pumpkin peels at 5–30 DAP were observed under a light microscope. The epicarp contains one epidermal layer and 16–20 layers of parenchymal cells with chloroplasts. At 5 DAP, there was no significant difference in the chloroplast distribution or chloroplast number of each parenchymal cell between '9-6' and 'U3-3-44'. Beginning 10 DAP, the chloroplasts of parenchymal cells near the epidermal layers in '9-6' began to clearly degrade during fruit development. The epicarp of the '9-6' mature fruit (at 30 DAP) was approximately 0.85 mm thick and was composed of 4–6 nonpigmented cell layers and 12–15 pigmented cell layers (Fig. S2a–d, Table S4). In the pigmented cell layers, there were 6–9 chloroplasts in each cell. Electron microscopy revealed that 90% of the chloroplasts in the nonpigmented cell layers were in various states of degradation. The chloroplast membranes were ruptured and exhibited vacuolation, causing a decrease in the chloroplast content. Thylakoids were also disorganized (Fig. 1e–g). However, in the 'U3-3-44' epicarp, the number of chloroplast in each parenchymal cell continued to increase during fruit development (Fig. S2e–h, Table S4). In the 'U3-3-44' mature fruit, the epicarp thickness was 1.15 mm, and there were 16–20 pigmented cell layers without nonpigmented cell layers. In the pigmented cell layers, there were 20–26 chloroplasts in each cell. Under an electron microscope, the structures of the chloroplast membranes maintained their integrity, the thylakoids structures were well organized, and full starch grains were observed (Fig. 1b–d). In the mature fruits, the cell volume did not significantly differ between '9-6' and 'U3-3-44'. Therefore, the difference between the epicarp cell layers of '9-6' and 'U3-3-44' was not the cell volume, but the presence of nonpigmented cell layers and the chloroplasts number of each cell in the pigmentary layer.

Inheritance of peel colour in *C. maxima*

Six generations of the '9-6' and 'U3-3-44' populations were constructed to analyse the inheritance of the peel colour phenotype. We collected peel colour data for all maternal fruits, and the phenotypic segregation of the population is presented in Table 1. All F_1, F_1', and BC_1P_2 ($F_1 \times$ U3-3-44) individuals had dark green peels, indicating the dominant nature of dark green peels in 'U3-3-44'. Among the BC_1P_1 ($F_1 \times$ 9-6) population, 26 individuals had dark green peels and 24 individuals had grey peels, which corresponded to a 1:1 dark green to grey peel segregation ratio ($\chi^2 = 0.08$, $p = 0.78$). Among the 187 F_2 individuals, 133 were dark green plants, and 54 were grey plants, which fit a 3:1 green to grey peel segregation ratio ($\chi^2 = 1.5$, $p = 0.22$). These data indicated that the dark green peel colour phenotype was controlled by a single dominant locus, which was named *CmaMg* (*mature green peel*).

Preliminary mapping of the *CmaMg* locus

A total of 40.18 and 43.09 million clean reads were obtained from the dark green pool and grey pool, respectively, which represented 39.27% and 39.34% of the total GC content, respectively, and the Q30 values reached 91.8% and 91.56%, respectively. The average read depths of the clean reads from the dark green and grey pools were 26× and 29×, respectively, and covered 97.91% and 98.01% of the genome, respectively. These results indicated that the sequencing results were reliable for gene mapping.

Preliminary BSA mapping for the dark green peel trait was obtained with the SNP-index and InDel-index between two gene pools by plotting the average SNP-index and InDel-index values, respectively. A total of 62,224 and 24,456 high-quality SNPs and InDels were identified between the two gene pools on all 20 chromosomes, respectively. The Δ(SNP-index) and Δ(InDel-index) were calculated and plotted by comparing the results of the SNP-index

Table 1 Segregation of peel colour in *C. maxima*

Population	Total	Green	Gray	Green:Gray	χ^{2a}	*P*
9-6 (P1)	20	–	20	–	–	–
U3-3-44 (P2)	20	20	–	–	–	–
F_1	20	20	–	–	–	–
F_1'	20	20	–	–	–	–
BC_1P_1	50	26	24	1:1	0.08	0.78
BC_1P_2	50	50	–	–	–	–
F_2	187	133	54	3:1	1.5	0.22

[a] $\chi^2(0.05,1) = 3.84$

and InDel-index of the two gene pools in the genomic positions. Regarding the intersection of association regions with Δ(SNP-index) and Δ(InDel-index) values, a 2.10-Mb region associated with the peel colour region was located at the start of chromosome 11, containing a total of 426 genes (Fig. S3). InDels were identified in the 2.10-Mb region based on the genome-wide recombination sequences of '9-6' and 'U3-3-44' for marker development, and 5 polymorphic InDel markers were developed to map the *CmaMg* locus. Genotype−phenotype joint analysis was performed using these InDel markers on the 177 F_2 individuals, and *CmaMg* was flanked by the markers M_85721 and M_535227. This locus spanned a genetic distance of approximately 2 cM and corresponded to a physical distance of approximately 449.51 kb (Fig. 3a).

Fine mapping of the *CmaMg* locus

Five new polymorphic InDel markers in the mapped region were developed and screened in 1,703 F_2 individuals for fine mapping. *CmaMg* was flanked by the closer markers M_336400 and M_535227, and the candidate region had a genetic distance of approximately 1.7 cM and corresponded to a physical distance of approximately 198.83 kb (Fig. 3b). A total of 21 and 7 recombinants were identified in this population by the markers M_336400 and M_535227, respectively (Table S5). F_3 offspring self-crossed from 28 recombinants were planted (25 plants from each recombinant), and the colour of the mature fruit on each plant was observed. Among the markers M_336400 and M_535227, 6 polymorphic InDel and KASP markers were used to genotype the 28 recombinant plants. Among the $F_{2:3}$ offspring of recombinant #53, there were dark green fruits and grey fruits, and the phenotype of F2-53 was identified as heterozygous. F2-53 originated from chromosome 'U3-3-44' at M_336400 to M_418696 and was heterozygous in the region from M_422692 to M_535227. The phenotype and genotype of recombinant F2-53 indicated a right margin of the target region on M_418696. Among the $F_{2:3}$ offspring of recombinant #626, all fruits were dark green peel, and the phenotype of F2-626 was identified as homozygous dominant genotypes. F2-626 originated from chromosome 'U3-3-44' at M_336400 to M_422692 and heterozygous in the region from M_454930 to M_535227. The phenotype and genotype of recombinant F2-626 indicated a left margin of the target region on M_454930. In the F_3 phenotype combined with the genotype of F_2 recombinants, markers M_418696 and M_454930 flanked the *CmaMg* locus at a physical distance of 36.23 kb, and M_422692 cosegregated with *CmaMg* (Fig. 3c, Table S5). F2-53 and F2-869 were the heterozygote genotypes and exhibited recombination between markers M_418696 and M_454930, so the two recombinants were important for gene mapping. F_3 and F_4 offspring (25 plants from each F_3 line) of the two recombinants were generated (Table S6). Among the F_4 offspring from F3-3, F3-8, and F3-9, all fruits were dark green peel, and the phenotype of F3-3, F3-8, and F3-9 was identified as homozygous dominant genotypes. Recombinants F3-3 and F3-9 were heterozygous in the region from M_336400 to M_422692, but homozygous for the dark green peel 'U3-3-44' allele in the region from M_454930 to M_535227. The recombinant F3-8 was homozygous for the grey peel '9-6' allele in the region from M_336400 to M422692, but homozygous for the dark green peel 'U3-3-44' allele in the region from M_454930 to M_535227. The phenotype and genotype of the recombinant F3-3, F3-8, and F3-9 stains indicated the right margin of the target region on M_422692. In conclusion, *CmaMg* flanked markers M_422692 and M_454930 with a physical distance of approximately 32.34 kb (Fig. 3d, Table S6).

To confirm the *CmaMg* candidate region via fine mapping, we generated 517 BC_1 individuals. Polymorphic markers in the region of M_336400 and M_535227 were screened in BC_1 individuals, and *CmaMg* was also flanked by the markers M_422692 and M_454930 through numerous recombinants (Fig. S4, Table S7). These results indicated that *CmaMg* was mapped to the region with a physical distance of approximately 32.34 kb between markers M_422692 and M_454930.

Predicted genes of *CmaMg*

According to the Cucurbit Genomics Database, there were five coding genes between markers M_422692 and M_454930: *CmaCh11G000880*, *CmaCh11G000890*, *CmaCh11G000900*, *CmaCh11G000910*, and *CmaCh11G000920*. The positions, InDels, SNPs, and predicted functions of these genes are summarized in Fig. 3 and Table 2. SNPs and InDels located in promoter elements, untranslated regions (UTRs), and CDS regions were analysed between '9-6' and 'U3-3-44'. Among the 5 genes, one gene (*CmaCh11G000900*) had nonsynonymous changes in the CDS region and 3 genes (*CmaCh11G000880*, *CmaCh11G000900*, and *CmaCh11G000910*) had changes in the promoter element.

In the *CmaCh11G000900* CDS, there were 1 nonsynonymous InDel and 2 nonsynonymous SNPs between the two parental lines. *CmaCh11G000900* (*CmaKNAT6*) encoded knotted1-like homeobox 6, which contained KNOX 1, KNOX 2, and ELK conserved domains and the homeobox KN domain (Fig. 4b). Earlier studies showed that *KNAT* genes influence the gradient of fruit chloroplast development through the regulation of downstream gene expression in tomato. *CmaKNAT6* of 'U3-3-44' contained 5 exons and 4 introns, and the first exon was 300 bp long. *CmaKNAT6* of '9-6' had a 580 bp deletion in part of the first exon and the

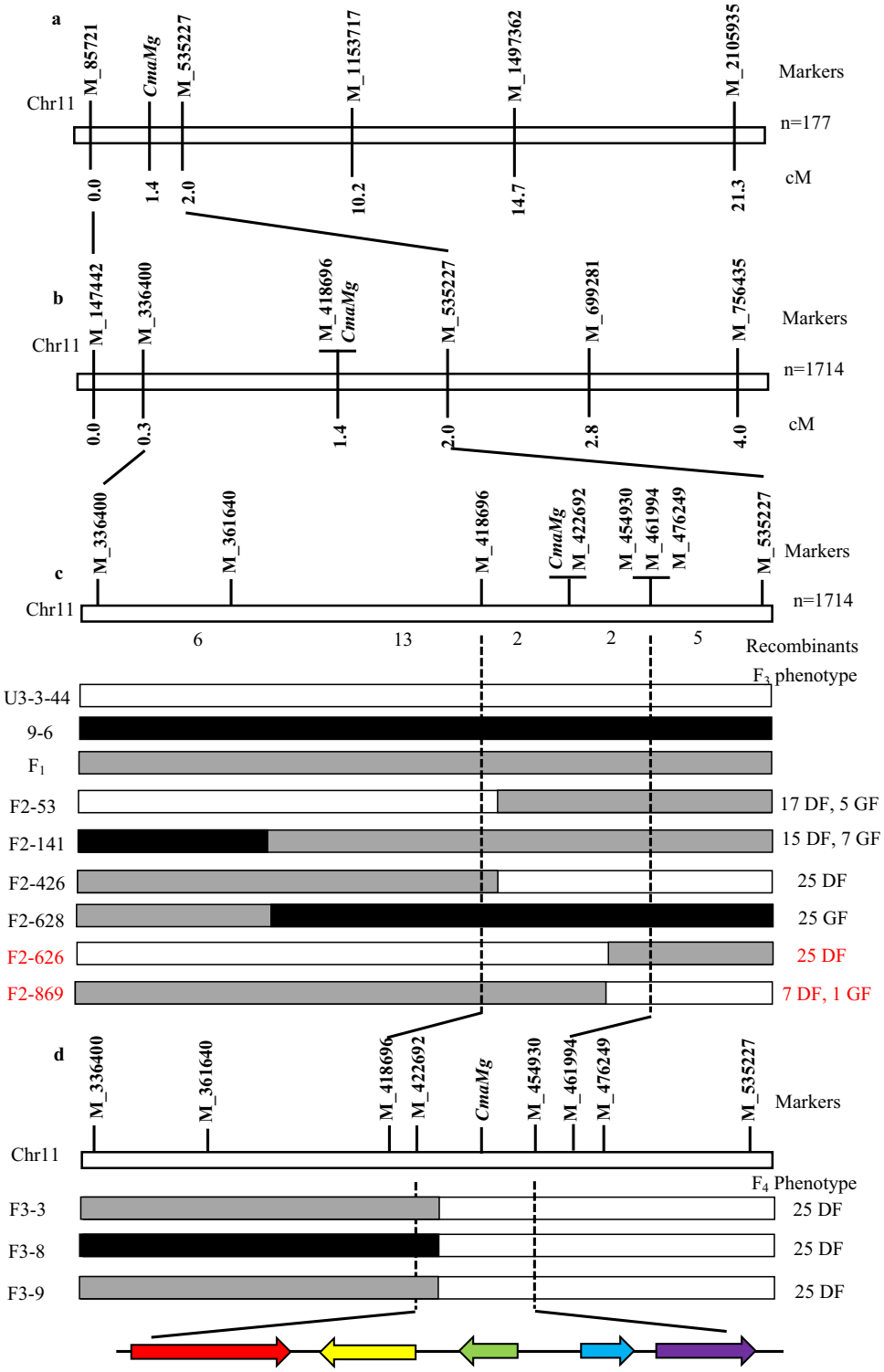

Fig. 3 Genetic and physical mapping of *CmaMg* using F_2, $F_{2:3}$, and $F_{3:4}$. **a** Genetic mapping of *CmaMg* by InDel markers. Using the 177 F_2 individuals from '9-6' × 'U3-3-44', *CmaMg* locus was located to a 449.51 kb region between markers M_85721 and M_535227 on chromosome 11. **b** and **c**. Fine mapping based on 1,714 F_2 plants and 28 recombinants with phenotype of $F_{2:3}$ offsprings defined the *CmaMg* locus to a 36.23 kb region between markers M_418696 and M_454930. DF indicates dark green fruit, GF indicates grey fruit. **d** Fine mapping based on F_3 genotype and $F_{3:4}$ phenotype defined the *CmaMg* locus to a 32.34 kb region between markers M_422692 and M_454930. The white colour codes homozygous 'U3-3-44', the black colour codes homozygous '9-6', and the grey colour codes heterozygous genotypes. The arrows represent genes, and the direction of the arrow indicates the direction of gene expression in the genome

first intron (195 bp in the first exon and 385 bp in the first intron). The InDel mutation of *CmaKNAT6* led to different sequences of the first exon (105 bp of the first exon and 15 bp of the first intron) and a premature stop codon (TAA). Therefore, CmaKNAT6 of '9-6' contained 40 amino acids without any conserved domains (Fig. 4).

To further identify the candidate genes of *CmaMg*, we measured expression levels of 5 genes at different stages (6 DBP to 40 DAP) and in different tissues (roots, stems, leaves, flowers, peels, and flesh) from '9-6' and 'U3-3-44'. As shown in Fig. 5, the expression of *CmaKNAT6* in '9-6' and 'U3-3-44' showed almost no change at 6 DBP to 5

Table 2 Predicted genes between marker M_422692 and marker M_454930

No	Gene ID	Location	Promoter element	UTR region	Intron region	CDS region	Type	Description
1	*CmaCh11G000880*	425970–436747	Yes	No	1 InDel	No	Syn	Modifier of SNC1 1-like protein
2	*CmaCh11G000890*	436783–437713	No	No	No	No	No	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B
3	*CmaCh11G000900*	437936–447202	Yes	No	1 InDel	1 InDel 2 SNPs	Nonsyn	Homeobox protein knotted-1-like 6 (KNOX)
4	*CmaCh11G000910*	449164–451690	Yes	No	No	No	No	WD repeat-containing protein 85
5	*CmaCh11G000920*	452326–460287	No	No	4 SNPs	No	Syn	Pleckstrin homology domain-containing family M member 3

Nonsyn: nonsynonymous, Syn: synonymous

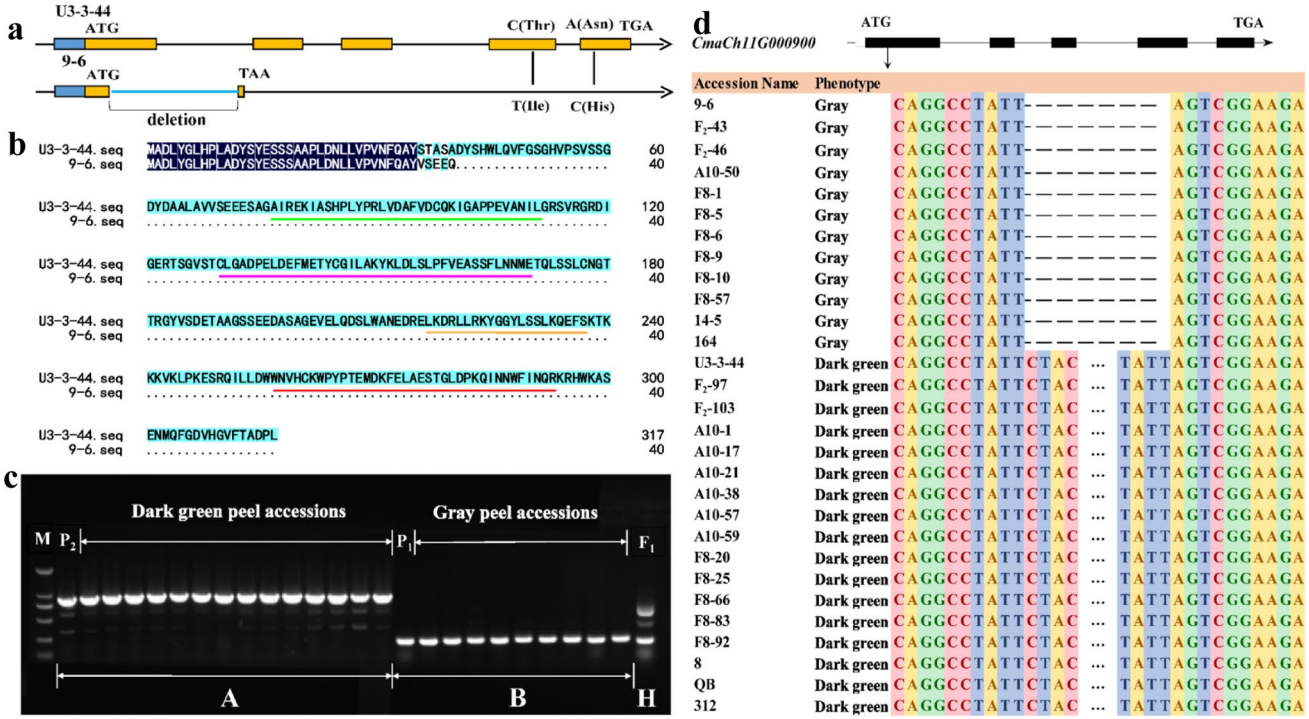

Fig. 4 Alignment of gene sequences and the predicted protein sequences of *CmaCh11G000900*. **a** Gene structure of '9-6' and 'U3-3-44'. **b** Protein sequences of *CmaCh11G000900* between '9-6' and 'U3-3-44'. The green line indicated KNOX 1 domain, the purple line indicated KNOX 2 domain, the orange line indicated ELK domain, and the red line indicated Homeobox KN domain. **c** InDel marker of germplasm resources. U3-3-44 (P2), 8, 10–17, 312, A10-1, A10-21, A10-38, A10-57, A10-59, F8-20, F8-25, F8-66, F8-83, F8-92, and QB are dark green peel accessions, and 9-6 (P1), 14–5, 164, A10-50, F8-1, F8-5, F8-6, F8-9, F8-10, and F8-57 are grey peel accessions. **d** Gene sequences of *CmaCh11G000900* in germplasm resources

DAP and then tended to sharply increase during the early fruit development stage while decreased gradually during the late period of fruit development. The expression of *CmaKNAT6* was greatest at 30 DAP in both '9-6' and 'U3-3-44'. The expression level of *CmaKNAT6* in grey peel '9-6' was significantly or extremely significantly greater than that in dark green peel 'U3-3-44' at the 5–40 DAP stage. Among the 5 genes, the expression levels of the other four genes did not show consistency with fruit development (Fig. 5). The tissue-specific expression patterns of the 5 genes in different tissues of pumpkin were analysed (Fig. S5). The 5 genes were expressed in almost all tissues. *CmaCh11G000880* and *CmaCh11G000920* were highly expressed in roots, flowers, and leaves; *CmaCh11G000890* was highly expressed in leaves; *CmaKNAT6* was highly

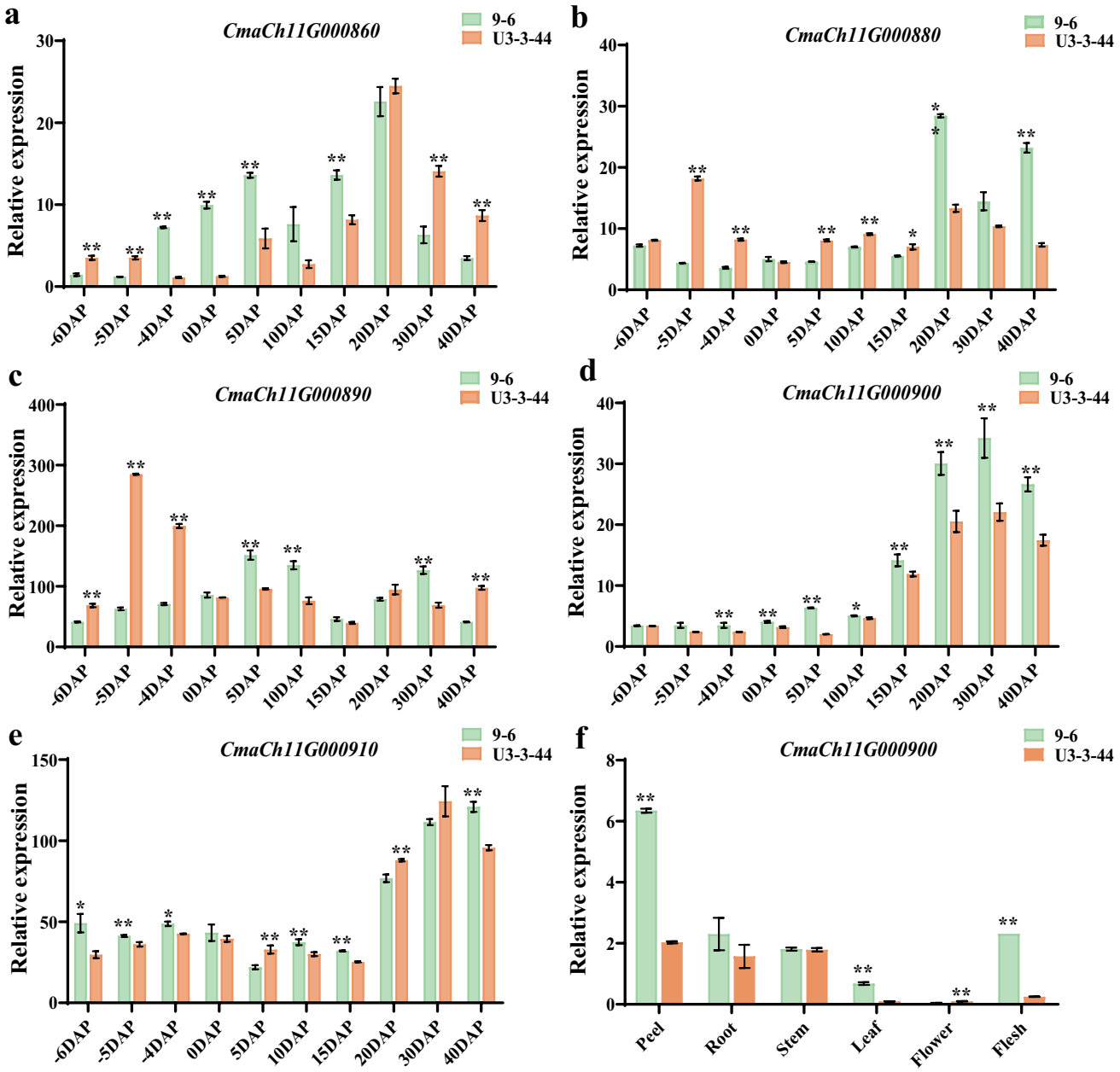

Fig. 5 Relative expression levels of five candidate genes during fruit development in '9-6' and 'U3-3-44'. **a** Expression levels of *CmaCh11G000880*. **b** Expression levels of *CmaCh11G000890*. **c** Expression levels of *CmaCh11G000900*. **d** Expression levels of *CmaCh11G000910*. **e** Expression levels of *CmaCh11G000920*. **f** The tissue-specific expression patterns of *CmaCh11G000900* in '9-6' and 'U3-3-44'. The relative expression levels of five genes were quantified using the $2^{-\Delta\Delta CT}$ method (Wang et al. 2022). The expression level of the *CmaCh11G000880* gene at 6 DBP in '9-6' was set to a value of 1 and used as a reference, respectively. Each was repeated three times, and three plants were mixed in equal amounts to form one replicate. ** indicates an extremely significant difference, $P < 0.01$; * indicates a significant difference, $P < 0.05$

expressed in peels; and *CmaCh11G000910* was highly expressed in stems, leaves, and flowers. Gene sequence analysis combined with gene expression analysis revealed that *CmaKNAT6* was the candidate gene associated with peel colour in *C. maxima*.

Allelic variation in the *CmaKNAT6* gene

To analyse the consistency between peel colour and allelic variations in *CmaKNAT6*, we collected *C. maxima* germplasms from 14 dark green peel accessions and 9 grey peel accessions. The full gene sequence of each accession of

CmaKNAT6 was obtained, and allelic diversity was analysed. 9 grey peel accessions had deletions at the 196th nucleotide in the first exon of *CmaKNAT6*, which was consistent with '9-6'. The *CmaKNAT6* sequence of 14 dark green peel accessions was consistent with that of 'U3-3-44' (Fig. 4d). An InDel marker was also constructed for a 580 bp deletion fragment. The bands of 14 dark green peel accessions were consistent with the 'U3-3-44', and those of 9 grey peel accessions were consistent with the '9-6' bands (Fig. 4c). These analyses indicated that the CDS variation in *CmaKNAT6* was associated with peel colour in *C. maxima* germplasms.

Phylogenetic analysis of *CmaKNAT6*

There are 21 genes in the KNAT family in the *C. maxima* genome. To characterize the evolutionary relationships among these genes and clarify the similarities in the KNAT genes among *Cucurbita* and other crops, we generated a phylogenetic tree of *KNAT* family genes from various plant species with protein sequences of *Arabidopsis*, *O.* sativa, *Z. mays*, *C. maxima*, *C. sativus*, *C. melo*, *C. lanatus*, and *S. lycopersicum* (Table S8). The 21 *KNAT* genes of *C. maxima* were divided into three groups, which was consistent with the *KNAT* gene classification of *Arabidopsis*. The colour-related proteins TKN2 and TKN4 in tomato and most *C. maxima* KNAT members, including the target gene *CmaKNAT6*, were clustered in class 1 (Fig. 6). *TKN2* and *TKN4* activate the expression of the *SlGLK2* and *APRR2*-like genes to increase chloroplast development and the distribution of chloroplasts in tomato fruits (Nadakuduti et al. 2014). *CmaKNAT6* was closely related to *TKN2* and *TKN4* in tomato, indicating that it may be involved in regulating chloroplast development in fruit peels and affecting peel colour.

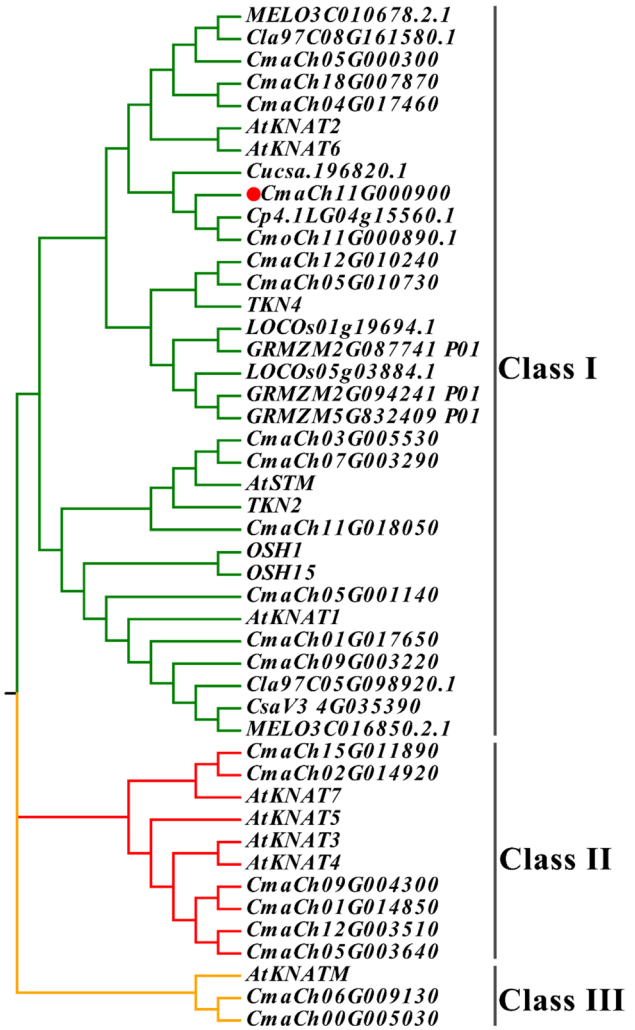

Fig. 6 Phylogenetic analysis of the candidate gene and the KNAT gene family in other species. The candidate gene *CmaMg* is marked in red. The numbers on nodes represent bootstrap values

Subcellular localization of *CmaKNAT6*

The subcellular localization o *CmaKNAT6* was determined, and the *35S*::CmaKNAT6-GFP fusion protein was transiently expressed in tobacco leaves. The green fluorescence signal of the empty vector was mainly concentrated in the nucleus and membrane under a laser confocal microscope, and for *35S*:CmaKNAT6-GFP, the green fluorescence signal was concentrated in the nucleus (Fig. S6).

RNA-seq analysis of dark green/grey fruit peels

Fruit peels of '9-6' and 'U3-3-44' were used as materials, and samples with the same peel colour period (6 DBP) and different peel colour period (20 DAP) were collected for RNA-seq analysis. Through sequencing, a total of 515.40 Mb of clean reads, with valid bases ranging from 96.22% to 97.17%, were obtained. In the clean reads of all 12 samples, the Q30 scores ranged from 93.89 to 94.60%, and the GC content ranged from 44.5 to 45% (Table S9), which demonstrated that high-quality sequencing results were obtained for further research. Three biological replicates of each sample type were clustered together using principal component analysis (PCA), and similar results were obtained through gene expression distance analysis with Pearson correlation (Fig. S7). The high correlation and similarity of expression profiles between replicates indicated high data reliability and reproducible replicate samples of RNA-seq.

To characterize the transcriptional variations associated with peel colour, we compared the transcripts between sample types. The genes that were differentially expressed in both G20 vs. DG20 and DG6 vs. DG20 were selected because they are related to fruit peel colour and were

subsequently subjected to Gene Ontology (GO) classification and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. The GO annotations of the differentially expressed genes (DEGs) are shown in Fig. 7a, and the main functional groups of the DEGs genes were related to 'photosynthesis', 'pyridine nucleotide metabolic process', 'nicotinamide nucleotide metabolic process', and 'oxidoreduction coenzyme metabolic process'. KEGG analysis revealed that the DEGs were mainly enriched in the 'carbon metabolism', 'carbon fixation in photosynthetic organisms', 'glyoxylate and dicarboxylate metabolism', 'photosynthesis', 'phenylpropanoid biosynthesis', and 'gingerol biosynthesis and flavonoid biosynthesis' pathways (Fig. 7b).

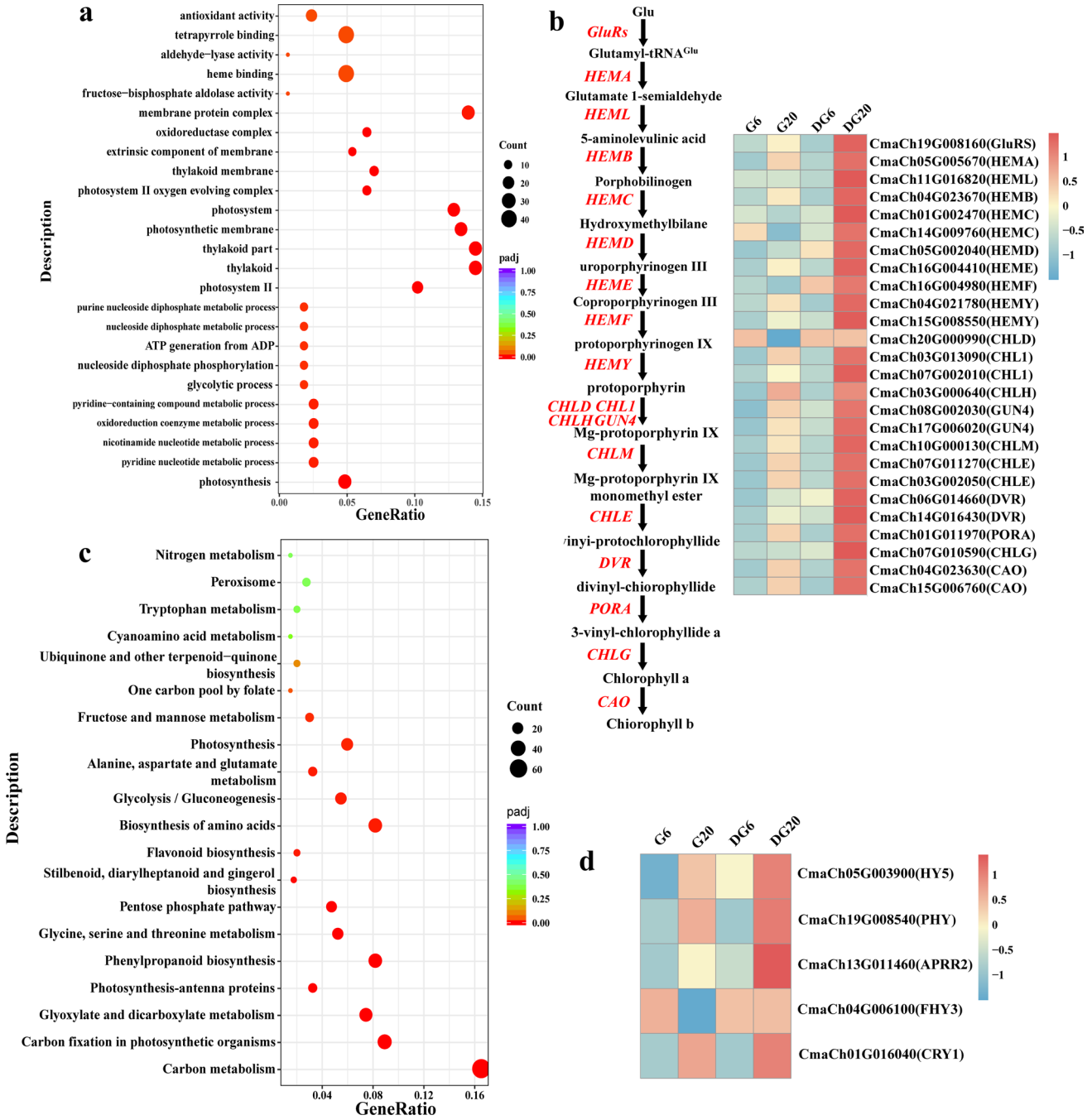

Fig. 7 Transcriptome profiling of fruit peel colour. **a** GO enrichment analysis of DEGs-related fruit peel colour. **b** KEGG enrichment analysis of DEGs related fruit peel colour. **c** DEGs related to chlorophyll biosynthesis pathway. **d** DEGs related to light signal transduction pathways and chloroplast functionality

DEGs are involved in chlorophyll biosynthesis, light signal transduction pathways, and chloroplast functionality

At different peel colour periods, the expression levels of genes related to 'photosynthesis' and 'carbon fixation in photosynthetic organisms' were greater in the dark green peel 'U3-3-44' than in the grey peel '9-6'. Chlorophyll biosynthesis pathway-related genes (*GluRS*, *HEMA*, *HEMB*, *HEMC*, *HEMD*, *HEME*, *HEMY*, *CHLD*, *CHL1*, *CHLH*, *GUN4*, *CHLM*, *CHLE*, *DVR*, *PORA*, *CHLG*, and *CAO*) were upregulated in 'U3-3-44' (Fig. 7c). Light signals activate the expression of genes related to chlorophyll synthesis, prompting plants to produce chlorophyll with biological functions under light conditions. This regulatory effect is dependent mainly on important transcription factors involved in the light signal transduction pathway. The positive regulatory factors directly involved in chlorophyll biosynthesis, *HY5* and *PHY*, were upregulated in 'U3-3-44'. Other transcription factors, such as *APRR2* and *far-red elongated hypocotyl 3* (*FHY3),* were upregulated in 'U3-3-44' (Fig. 7d). *Cryptochrome1* (*CRY1*), a coregulatory gene of chlorophyll biosynthesis with chloroplast functionality, was also upregulated in 'U3-3-44'. The data indicated that the *CmaKNAT6* gene may regulate chlorophyll accumulation by influencing the expression of transcription factors related to light signal transduction and chloroplast functionality.

Discussion

Fruit peel colour is an important quality trait in Cucurbitaceae crops and directly influences consumer choice. *C. maxima* has a variety of peel colours, such as red, orange, yellow, green, and grey. Therefore, *C. maxima* is a valuable material for research on this mechanism. For flesh-used *C. maxima*, red, green, and grey were the dominant colours of the fruit peel. In the present study, we used '9-6', which has a grey peel and 'U3-3-44', which has a dark green peel, as materials to identify the genes controlling green peel colour in *C. maxima*. The dominant gene *CmaMg*, which regulates peel colour, was narrowed down to a 32.34-kb region on chromosome 11 by map-based cloning, and *CmaKNAT6* was identified as a promising candidate gene for *CmaMg* according to the reported function, sequence alignment, and expression analyses. This gene not only provides a new peel colour gene resource but also provides an important basis for understanding the regulatory mechanism of peel colour in Cucurbitaceae plants. The InDel marker in the *CmaMg* coding region could identify grey/dark green fruit peel colour at the seedling stage and could be used in marker-assisted selection breeding for fruit peel colour in *C. maxima*.

During fruit development, the peel colour of '9-6' was light green and gradually turned grey, and the peel colour of 'U3-3-44' changed from light green to dark green. Microscopic observation of the dark green and grey pericarps revealed that the epidermal cell layers of '9-6' and 'U3-3-44' differed in the number of chloroplasts and the structural integrity of the chloroplasts. This phenomenon indicated that *CmaMg* influenced chloroplast development and chloroplast synthesis in the fruit pericarp. F_2 populations of '9-6' and 'U3-3-44' were constructed, and the peel colour phenotype was observed. In the F_2 population, grey peeled fruits and green peeled fruits were easy to distinguish with the naked eye, while the fruits in the green-peeled group exhibited varying degrees of green peel colour. These findings indicated that *CmaMg* might determine green peel colour, and other genes might determine the depth of green colour.

Chlorophyll and carotenoids are the main pigments determining peel colour in Cucurbitaceae plants (Duan et al. 2017). Importantly, the ratio of carotenoids to chlorophyll determined the plant colour. In young cucumber fruit, the carotenoid/chlorophyll ratio of green, reseda, blocky-green, and bottle green peels was lower than or equal to 1, and the carotenoid/chlorophyll ratio of white peels was greater than 1 (Sun et al. 2003). It was indicated that chlorophyll can affect the colour of carotenoids when the carotenoid/chlorophyll ratio is lower than 1. In this study, the chlorophyll a, chlorophyll b, total chlorophyll, and carotenoid contents of the 'U3-3-44' fruit pericarp were significantly greater than those of the '9-6' fruit pericarp during fruit development. These findings indicated that *CmaMg* affected not only chlorophyll content but also carotenoid accumulation. In the 'U3-3-44' fruit pericarp, the total chlorophyll content was significantly greater than the carotenoid content throughout the fruit development stage, and the carotenoid/chlorophyll ratio ranged from 0.1 to 0.25; therefore, the peel colour of 'U3-3-44' was green. The chlorophyll and carotenoid contents of the stems and leaves were also determined, and the chlorophyll and carotenoid contents in 'U3-3-44' were significantly greater than those in '9-6'. Further analysis of tissue expression showed that *CmaMg* was expressed not only in the fruit peel but also in the stem and leaf. These findings indicated that *CmaMg* regulated pigment accumulation in multiple organs of *C. maxima*.

KNAT genes influence multiple aspects of plant morphology and are usually expressed in apical meristems, where they maintain the pluripotent cell populations required for organ initiation (Belles-Boix et al. 2006; Hay and Tsiantis 2010; Nadakuduti et al. 2014). The *KNAT* gene not only affects shoot apical meristem formation and leaf development but also influences plant height disrupts apical dominance and alters patterns of cell division and elongation (Douglas et al. 2002; Long et al. 1996; Venglat et al. 2002). The *Arabidopsis KNAT* genes,

SHOOT MERISTEMLESS (*STM*), *BREVIPEDICELLUS* (*BP*), *KNAT2*, and *KNAT6* influence the development of the inflorescence and carpels together with the architecture of abscission zones (Ragni et al. 2008; Scofield et al. 2007; Shi et al. 2011). *KNAT* genes may play an active role in the effects of plant hormone cytokines on chloroplast biogenesis and development (Cortleven and Schmülling 2015; Nadakuduti et al. 2014; Tanaka et al. 2011). Class I *KNAT* genes usually affect plant morphology by maintaining hyphal tissue activity (Hay and Tsiantis 2010; Parnis et al. 1997; Venglat et al. 2002). The role of *KNAT* genes (*TKN2* and *TKN4*) in regulating fruit chloroplast development has been demonstrated in the *Curl* (*Cu*) mutants. In this study, *CmaKNAT6*, *STM, TKN2*, and *TKN4* were grouped into the same class, and *CmaKNAT6* might perform similar functions in chloroplast development. *CmaKNAT6* contains the KNOX 1, KNOX 2, and ELK conserved domains and the homeobox KN domain. The KNOX1 domain plays a role in suppressing target gene expression, the KNOX2 domain is necessary for homodimerization and essential for gene function, and the homeobox KN domain is a DNA-binding factor that is involved in the transcriptional regulation of key developmental processes. In '9-6', the InDel mutation of *CmaKNAT6* causes premature termination of transcription, and the four conserved domains disappear. Due to the absence of important domains, *CmaKNAT6* likely cannot bind to other transcription factors to perform its normal regulatory function in chloroplast development. Interestingly, the gene expression of *CmaKNAT6* in '9-6' was significantly greater than that in 'U3-3-44' during fruit development. We speculated that nonfunctional *CmaKNAT6* cannot maintain chloroplast development or chloroplast biosynthesis in fruit peels and that plants need to maintain their basic functions through feedback regulation, resulting in increased expression of the *CmaKNAT6* gene in '9-6'.

TKN2 and *TKN4* have been shown to act upstream of *SlGLK2* and the related gene *ARABIDOPSIS PSEUDO RESPONSE REGULATOR 2-LIKE* (*SlAPRR2-LIKE*) to establish a latitudinal gradient in their expression in developing fruit, leading to a chloroplast developmental gradient (Nadakuduti et al. 2014). The *GLK* gene family is involved in the transcriptional regulation of chloroplast biogenesis in diverse groups of plants, and *GLK* can promote chloroplast development in nonphotosynthetic organs such as roots and fruits. APRR2 was identified as an important transcription factor that regulates chlorophyll accumulation in *C. sativus* (Liu et al. 2016), *C. melo*, *C. lanatus* (Oren et al. 2019; Zhao et al. 2019), and *C. pepo* (Ding et al. 2024). In our study, *APRR2* was upregulated at 20 DAP in 'U3-3-44'. *GLK* was expressed at low levels in all transcriptome samples, and there was no difference in the expression between the green peel and grey peel samples. These findings indicated that the expression of *APRR2* may be regulated by *CmaKNAT6*, but whether *CmaKNAT6* regulates downstream *APRR2* needs further verification.

Transcription factors involved in light signal transduction and chloroplast functionality were selected. HY5 is a central transcription factor that regulates many light-associated events including photomorphogenesis and chloroplast development downstream of photoreceptors, and many chlorophyll biosynthesis genes are direct targets of HY5 (Lee et al. 2007). The transposase-derived transcription factor FHY3 is a positive regulator of PHYA signalling and functions in diverse developmental processes, including chlorophyll biosynthesis (Wang and Wang 2015). *HEMA1* was identified as a putative direct target of FHY3, and the expression of *GUN4* and *CHLH* was also regulated by FHY3 (Ouyang et al. 2011). FHY3 physically interacts with HY5 through its DNA binding domain (Li et al. 2010), which affects PHYA signalling and the circadian clock (Wang and Wang 2015). Plants need to coordinate chlorophyll biosynthesis with the formation of photosynthetic machinery during development to prevent photooxidative damage to chloroplasts. HY5 represses its target genes via CRY1 when chloroplasts are dysfunctional (Ruckle et al. 2007), and loss-of-function mutations in CRY1 or HY5 cause strong photodamage under light conditions (Ruckle et al. 2007). CRY1-HY5-mediated light signalling is crucial for controlling chloroplast functionality and the development of the photosynthetic machinery according to light conditions. In our study, the central gene *HY5* and its regulated genes *FHY3* and *CRY1* were upregulated at 20 DAP in 'U3-3-44'. Based on transcriptomic data and morphological observations of the fruit pericarp, we deduced that CmaKNAT6 upregulated the expression of *HY5*, *PHT*, and *FHY3* to promote chlorophyll biosynthesis and upregulated the expression of *HY5 and CRY1* to maintain chloroplast functionality under light conditions. Therefore, the fruit peel colour of 'U3-3-44' gradually turned dark green during fruit development. Nonfunctional CmaKNAT6 could no longer prevent photooxidative damage to chloroplasts, causing disorganized thylakoids and incomplete chloroplasts in '9-6'. During fruit development, the presence of nonpigmented cell layers gradually increased, and the fruit peel of '9-6' gradually turned grey. This study provides a theoretical basis for further evaluating the regulatory mechanism of pumpkin peel colour-related genes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-024-04741-7.

Acknowledgements This work was supported by Grants from the National Natural Science Foundation of China (32072590, 32272723, 31872124).

Author contribution statement YW, the corresponding author, analysed the data and oversaw manuscript development. SQ, the

corresponding author, oversaw all activities related to the project implementation and manuscript development. CW and WD analysed the data and prepared a draft of the manuscript. FC performed the fine mapping of the candidate gene. KZ performed gene cloning. YH participated in the experiments. GW and WX provided germplasm resources planting materials. All authors reviewed and approved this submission.

Funding This study was fund by the National Natural Science Foundation of China (32072590, 32272723, 31872124).

Data availability The raw datasets of BSA-seq generated during the current study are available in the NCBI repository, BioProject PRJNA1128793 and PRJNA1120204.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

Ethical approval The experiments comply with the current laws of the country in which we were performed**.**

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Ballester AR, Molthoff J, de Vos R et al (2009) Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. Plant Physiol 152:71–84

Barros L, Dueñas M, Pinela J et al (2012) Characterization and quantification of Phenolic compounds in four tomato (*Lycopersicon esculentum* L.) farmers' varieties in Northeastern portugal homegardens. Plant Foods Hum Nutr 67:229–234

Barry CS, McQuinn RP, Chung MY et al (2008) Amino acid substitutions in homologs of the STAY-GREEN protein are responsible for the *green-flesh* and *chlorophyll retainer* mutations of tomato and pepper. Plant Physiol 147:179–187

Belles-Boix E, Hamant O, Witiak SM, Morin H, Traas J, Pautot V (2006) KNAT6: an *Arabidopsis* homeobox gene involved in meristem activity and organ separation. Plant Cell 18:1900–1907

Chayut N, Yuan H, Ohali S, Meir A, Sa'ar U, Tzuri G, Zheng Y, Mazourek M, Gepstein S, Zhou X, Portnoy V, Lewinsohn E, Schaffer AA, Katzir N, Fei Z, Welsch R, Li L, Burger J, Tadmor Y (2017) Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. Plant Physiol 173:376–389

Chomicki G, Schaefer H, Renner SS (2020) Origin and domestication of *Cucurbitaceae* crops: insights from phylogenies, genomics and archaeology. New Phytol 226:1240–1255

Cortleven A, Schmülling T (2015) Regulation of chloroplast development and function by cytokinin. J Exp Bot 66:4999–5013

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, Williams C, Reich M, Winckler W, Getz G (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28:1530–1532

Ding W, Wang Y, Qi C, Luo Y, Wang C, Xu W, Qu S (2021) Fine mapping identified the gibberellin 2-oxidase gene CpDw leading to a dwarf phenotype in squash (*Cucurbita pepo* L.). Plant Sci 306:110857

Ding W, Luo Y, Li W, Chen F, Wang C, Xu W, Wang Y, Qu S (2024) Fine mapping and transcriptome profiling reveal *CpAPRR2* to modulate immature fruit rind color formation in zucchini (*Cucurbita pepo*). Theor Appl Genet 137(7):167

Douglas SJ, Chuck G, Dengler RE, Pelecanda L, Riggs CD (2002) KNAT1 and ERECTA regulate inflorescence architecture in *Arabidopsis*. Plant Cell 14:547–558

Duan Y, Xiang C, Liu X, Ma W, Sun T, Wang C (2017) Effect of rind structure and pigment composition on rind color in *Cucurbita maxima* (in Chinese). China Veg 11:33–39

Ezin V, Gbemenou UH, Ahanchede A (2022) Characterization of cultivated pumpkin (*Cucurbita moschata* Duchesne) landraces for genotypic variance, heritability and agromorphological traits. Saudi J Biol Sci 29:3661–3674

Feder A, Burger J, Gao S, Lewinsohn E, Katzir N, Schaffer AA, Meir A, Davidovich-Rikanati R, Portnoy V, Gal-On A, Fei Z, Kashi Y, Tadmor Y (2015) A kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in muskmelon. Plant Physiol 169:1714–1726

Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, Saitoh H, Mitsuoka C, Utsushi H, Uemura A, Kanzaki E, Kosugi S, Yoshida K, Cano L, Kamoun S, Terauchi R (2013) MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS ONE 8:1–10

Fernandez-Moreno JP, Tzfadia O, Forment J et al (2016) Characterization of a new pink-fruited tomato mutant results in the identification of a null allele of the SlMYB12 transcription factor. Plant Physiol 171:1821–1836

Ge Y, Li X, Yang XX, Cui CS, Qu SP (2015) Short communication genetic linkage map of *Cucurbita maxima* with molecular and morphological markers. Genet Mol Res 14:5480–5484

Hao N, Du Y, Li H, Wang C, Wang C, Gong S, Zhou S, Wu T (2018) CsMYB36 is involved in the formation of yellow green peel in cucumber (*Cucumis sativus* L.). Theor Appl Genet 131:1659–1669

Hay A, Tsiantis M (2010) KNOX genes: versatile regulators of plant development and diversity. Development 137:3153–3165

Hill JT, Demarest BL, Bisgrove BW (2013) MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res 23:687–697

Hwang I, Kim Y, Han J, Nou IS (2016) Orange color is associated with CYC-B expression in tomato fleshy fruit. Mol Breeding 36:42

Inskeep WP, Bloom PR (1985) Extinction coefficients of chlorophyll a and b in N, N-dimethylformamide and 80% acetone. Plant Physiol 77:483–485

Jeong H-B, Jang S-J, Kang MY et al (2020) Candidate gene analysis reveals that the fruit color locus *C1* corresponds to PRR2 in pepper (*Capsicum frutescens*). Front Plant 11:399

Kang SI, Rahim MA, Afrin KS et al (2018) Expression of anthocyanin biosynthesis-related genes reflects the peel color in purple tomato. Hortic Environ Biotechnol 59:435–445

Kiramana JK, Isutsa DK (2017) First detailed morphological characterisation of qualitative traits of extensive naturalized pumpkin germplasm in Kenya. Int J Dev Sci 6:500–525

Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y et al (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell 19:731–749

Li J, Li G, Gao S, Martinez C, He G, Zhou Z et al (2010) *Arabidopsis* transcription factor ELONGATED HYPOCOTYL5 plays a role in the feedback regulation of phytochrome A signaling. Plant Cell 22:3634–3649

Li Y, Wen C, 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 126:2187–2196

Li B, Zhao S, Dou J, Ali A, Gebremeskel H, Gao L, He N, Lu X, Liu W (2019) Genetic mapping and development of molecular

markers for a candidate gene locus controlling rind color in watermelon. Theor Appl Genet 132:2741–2753
Liu H, Meng H, Pan Y, Liang X, Jiao J, Li Y, Chen S, Cheng Z (2015) Fine genetic mapping of the white immature fruit color gene w to a 33.0-kb region in cucumber (*Cucumis sativus* L.). Theor Appl Genet 128:2375–2385
Liu H, Jiao J, Liang X, Liu J, Meng H, Chen S, Li Y, Cheng Z (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
Liu D, Yang H, Yuan Y, Zhu H, Zhang M, Wei X, Sun D, Wang X, Yang S, Yang L (2020) Comparative transcriptome analysis provides insights into yellow rind formation and preliminary mapping of the *Clyr* (*yellow rind*) gene in watermelon. Front Plant Sci 11:192
Long JA, Moan EI, Medford JI, Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature 379:66–69
Lun Y, Wang X, Zhang C, Yang L, Gao D, Chen H, Huang S (2016) A *CsYcf54* variant conferring light green coloration in cucumber. Euphytica 208:509–517
Luo F, Cheng S-C, Cai J-H et al (2019) Chlorophyll degradation and carotenoid biosynthetic pathways: gene expression and pigment content in broccoli during yellowing. Food Chem 297:124964
Lux A, Morita S, Abe J, Ito K (2005) An improved method for clearing and staining free-hand sections and whole-mount samples. Ann Bot 96:989–996
Ma L, Liu Z, Cheng Z, Gou J, Chen J, Yu W, Wang P (2021) Identification and application of *BhAPRR2* controlling peel colour in wax gourd (*Benincasa hispida*). Front Plant Sci 12:716772
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628
Nadakuduti SS, Holdsworth WL, Klein CL, Barry CS (2014) *KNOX* genes influence a gradient of fruit chloroplast development through regulation of *GOLDEN2-LIKE* expression in tomato. Plant J 78:1022–1033
Nguyen CV, Vrebalov JT, Gapper NE et al (2014) Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. Plant Cell 26:585–601
Niu J, Chen Q, Lu X, Wang X, Tang Z, Liu Q, Lei F, Xu X (2023) Fine mapping and identifying candidate gene of *Y* underlying yellow peel in *Cucurbita pepo*. Front Plant Sci 14:1159937
Oren E, Tzuri G, Vexler L, Dafna A, Meir A, Faigenboim A, Kenigswald M, Portnoy V, Schaffer AA, Levi A, Buckler ES, Katzir N, Burger J, Tadmor Y, Gur A (2019) The multi-allelic *APRR2* gene is associated with fruit pigment accumulation in melon and watermelon. J Exp Bot 70:3781–3794
Ouyang X, Li J, Li G, Li B, Chen B, Shen H et al (2011) Genome-wide binding site analysis of FAR-RED ELONGATED HYPOCOTYL3 reveals its novel function in *Arabidopsis* development. Plant Cell 23:2514–2535
Parnis A, Cohen O, Gutfinger T, Hareven D, Zamir D, Lifschitz E (1997) The dominant developmental mutants of tomato, Mouse-ear and Curl, are associated with distinct modes of abnormal transcriptional regulation of a Knotted gene. Plant Cell 9:2143–2158
Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667
Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8–15
Ragni L, Belles-Boix E, Günl M, Pautot V (2008) Interaction of KNAT6 and KNAT2 with BREVIPEDICELLUS and PENNY-WISE in *Arabidopsis* inflorescences. Plant Cell 20:888–900
Ruckle ME, DeMarco SM, Larkin RM (2007) Plastid signals remodel light signaling networks and are essential for efficient chloroplast biogenesis in *Arabidopsis*. Plant Cell 19:3944–3960
Salamov AA, Solovyev VV (2000) Ab initio gene finding in drosophila genomic DNA. Genome Res 10(4):516
Scofield S, Dewitte W, Murray JA (2007) The *KNO*X gene *SHOOT MERISTEMLESS* is required for the development of reproductive meristematic tissues in *Arabidopsis*. Plant J 50:767–781
Shi CL, Stenvik GE, Vie AK, Bones AM, Pautot V, Proveniers M, Aalen RB, Butenko MA (2011) Arabidopsis class I KNOTTED-like homeobox proteins act downstream in the IDA-HAE/HSL2 floral abscission signaling pathway. Plant Cell 23:2553–2567
Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1:2019–2025
Sun X, Wang B, Gu S, Wang Z (2003) Correlations of immature skin color and pigments in cucumber. Hortic Plant J 6:721
Sun X, Xie F, Chen Y, Guo Z, Dong L, Qin L, Shi Z, Xiong L, Yuan R, Deng W, Jiang Y (2022) Glutamine synthetase gene *PpGS1.1* negatively regulates the powdery mildew resistance in Kentucky bluegrass. Hortic Res 9:196
Tadmor Y, Burger J, Yaakov I, Feder A, Libhaber SE, Portnoy V, Meir A, Tzuri G, Sa'ar U, Rogachev I, Aharoni A, Abeliovich H, Schaffer AA, Lewinsohn E, Katzir N (2010) Genetics of flavonoid, carotenoid, and chlorophyll pigments in melon fruit rinds. J Agric Food Chem 58:10722–10728
Tanaka R, Kobayashi K, Masuda T (2011) Tetrapyrrole Metabolism in *Arabidopsis thaliana*. Arabidopsis Book 9:145
Venglat SP, Dumonceaux T, Rozwadowski K, Parnell L, Babic V, Keller W, Martienssen R, Selvaraj G, Datla R (2002) The homeobox gene *BREVIPEDICELLUS* is a key regulator of inflorescence architecture in *Arabidopsis*. Proc Natl Acad Sci U S A 99:4730–4735
Wang H, Wang H (2015) Multifaceted roles of FHY3 and FAR1 in light signaling and beyond. Trends Plant Sci 20:453–461
Wang C, Li W, Chen F, Cheng Y, Huang X, Zou B, Wang Y, Xu W, Qu S (2022) Genome-wide identification and characterization of members of the ACS gene family in *Cucurbita maxima* and their transcriptional responses to the specific treatments. Int J Mol Sci 23:8476
Xanthopoulou A, Montero-Pau J, Mellidou I, Kissoudis C, Blanca J, Picó B, Tsaballa A, Tsaliki E, Dalakouras A, Paris HS, Ganopoulou M, Moysiadis T, Osathanunkul M, Tsaftaris A, Madesis P, Kalivas A, Ganopoulos I (2019) Whole-genome resequencing of *Cucurbita pepo* morphotypes to discover genomic variants associated with morphology and horticulturally valuable traits. Hortic Res 6:94
Yoo H, Park W, Lee G-M et al (2017) Inferring the genetic determinants of fruit colors in tomato by carotenoid profiling. Molecules 22:764
Zhao G, Lian Q, Zhang Z, Fu Q, He Y, Ma S, Ruggieri V, Monforte AJ, Wang P, Julca I, Wang H, Liu J, Xu Y, Wang R, Ji J, Xu Z, Kong W, Zhong Y, Shang J, Pereira L, Argyris J, Zhang J, Mayobre C, Pujol M, Oren E, Ou D, Wang J, Sun D, Zhao S, Zhu Y, Li N, Katzir N, Gur A, Dogimont C, Schaefer H, Fan W, Bendahmane A, Fei Z, Pitrat M, Gabaldón T, Lin T, Garcia-Mas J, Xu Y, Huang S (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat Genet 51:1607–1615

Zheng Y, Wu S, Bai Y, Sun H, Jiao C, Guo S, Zhao K, Blanca J, Zhang Z, Huang S, Xu Y, Weng Y, Mazourek M, Reddy UK, Ando K, McCreight JD, Schaffer AA, Burger J, Tadmor Y, Katzir N, Tang X, Liu Y, Giovannoni JJ, Ling KS, Wechter WP, Levi A, Garcia-Mas J, Grumet R, Fei Z (2019) Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. Nucleic Acids Res 47(D1):D1128–D1136

Zhong YJ, Zhou YY, Li JX, Yu T, Wu TQ, Luo JN, Luo SB, Huang HX (2017) A high-density linkage map and QTL mapping of fruit-related traits in pumpkin (*Cucurbita* moschata Duch.). Sci Rep 7:12785

Zhu L, Wang Y, Zhang Z, Hu D, Wang Z, Hu J, Ma C, Yang L, Sun S, Li Y (2022) Chromosomal fragment deletion in *APRR2*-repeated locus modulates the dark stem color in *Cucurbita pepo*. Theor Appl Genet 135:4277–4288