Genome-wide Association Analysis Reveals a Novel QTL CsFS1 for Fruit Skin Color in Cucumber

Xiaolin Geng (gengxl@cau.edu.cn)
China Agricultural University College of Horticulture
Hongyu Huang
Tianjin Kernel Cucumber Research Institute
Qinqin Yang
China Agricultural University College of Horticulture
Lidong Zhang
Tianjin Kernel Cucumber Research Institute
Weiliang Kong
Tianjin Kernel Cucumber Research Institute
Huizhe Wang
Tianjin Academy of Agricultural Sciences
Aimin Wei
Tianjin Academy of Agricultural Sciences
Shengli Du
Tianjin Academy of Agricultural Sciences
Ruihuan Yang
Tianjin Academy of Agricultural Sciences
Jiawang Li
Tianjin Kernel Cucumber Research Institute
Tao Lin
China Agricultural University College of Horticulture
Yuhe Li
Tianjin Academy of Agricultural Sciences

Research Article

Keywords: Cucumber, fruit skin color, BSA-seq, GWAS, GATA transcription factor

Posted Date: January 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-578494/v2

Abstract

Background: Cucumber is an important melon crop in the world, with different skin colors. However, the candidate genes and the underlying genetic mechanism for such an important trait in cucumber are unknown. In this study, a locus controlling fruit skin color was found on chromosome 3 of cucumber genome.

Results: In this study, the light green inbred line G35 and the dark green inbred line Q51 were used as parents to construct an F_2 population. On chromosome 3, we identified a major QTL *CsFS1* (Fruit skin 1). We further narrowed down the *CsFS1* locus to a 94-kb interval containing 15 candidate genes. Among these genes, *Csa3G912920*, which encodes a GATA transcription factor, was expressed at a higher level in the pericarp of the NIL-1334 line (with light-green fruit skin) than in that of the NIL-1325 line (with dark-green fruit skin). This study provides a new allele for the improvement of cucumber fruit skin color.

Conclusion: A major QTL that controls fruit skin color in cucumber, *CsFS1*, was identified in a 94-kb region that harbors the strong candidate gene *CsGATA1*.

Background

Fruit skin color is a valuable trait in the horticulture industry because it strongly influences consumer preference and exhibits extensive phenotypic variation that can be used in breeding. Many quantitative trait loci (QTLs) and genes related to fruit skin color have been detected and/or cloned in crops. In melon, skin color is determined by the composition and content of pigments such as carotenes, flavonoids, and chlorophylls [1]. In yellow casaba muskmelon, *CmKFB*, which encodes a kelch domain-containing F-box protein, was identified on chromosome 10 and shown to downregulate the accumulation of naringenin chalcone [2]. *MEL03C003375* on chromosome 4 and *MELO3C003097* on chromosome 8 were also shown to be closely associated with skin color [3]. In watermelon, *qrc-c8-1* on chromosome 8 controls the green shade of fruit skin; it was identified by high-density genetic mapping of recombinant inbred lines and explained 49.942% of the phenotypic variation in skin color [4]. *Cla002755* and *Cla002769* on chromosome 4 are markers for yellow skin and were identified by bulked segregant analysis sequencing (BSA-seq) and genome-wide association studies (GWAS) [5]. In tomato, *SlMYB12* was mapped to chromosome 1; it corresponded to the pink gene *y* and controlled the accumulation of yellow-colored flavonoids in the tomato fruit epidermis [6, 7]. In pepper, three independent pairs of genes (*y*, *c1*, and *c2*) and two QTLs (*pc8.1* and *pc10*) were identified as controlling ripe fruit color and chlorophyll content [8].

Cucumber (*Cucumis sativus* L., 2n=2x=14) is an economically important cucurbitaceous crop worldwide, with a total global production of 75.2 million tons, of which 56.2 million tons (74.7%) were produced throughout the Chinese mainland in 2018 (data available at http://www.fao.org/). The skin color of cucumber fruit is an important agronomic character that affects consumer choice. The locus *w* that controls the white fruit skin of cucumber was mapped to an 8.2-kb region on chromosome 3 between the LH39253 and ASPCR39250 markers and contains only one gene, *Csa3G904140* (*APRR2*) [9]. *APRR2*

encodes a nuclear localization transcription factor and controls fruit skin color by reducing the content of chlorophyll and chloroplasts [10, 11]. Cucumber *Csa7G051430* was identified by BSA-seq of extreme-phenotype F_2 individuals from a cross between the light-green skin mutant *lgp* and the wild type 406. It is homologous to *Arabidopsis ARC5*, which plays an important role in chloroplast division [12, 13]. Similarly, *Csa6G133820*, mapped through the light-green leaf and fruit skin mutant M218, encodes a Ycf54-like protein required for chlorophyll synthesis named *CsYcf54* [14, 15]. *Csa2G352940* (*CsMYB36*), encoding the transcription factor MYB36, regulates yellow-green peel color in cucumber [16]. To date, the mechanism that controls green fruit skin color in cucumber remains unclear. Further study of skin color inheritance and identification of candidate genes associated with green skin color will therefore provide valuable information.

BSA-seq and GWAS are simple and effective methods for the identification of molecular markers associated with target genes and QTLs that control traits of interest [17, 18]. This study was designed to determine the inheritance pattern of green fruit skin color and to map major skin color QTLs. BSA-seq analysis detected a genomic region harboring a major fruit skin color QTL, *CsFS1*, on chromosome 3, and it was further validated by GWAS analysis. This study also provides preliminary evidence that *Csa3G912920* is the probable candidate gene in the *CsFS1* locus.

Results

Phenotypic analysis of fruit skin color in cucumber

The inbred lines G35 (light-green cucumber) and Q51 (dark-green cucumber) were used as parents for fine mapping of fruit skin color. The fruit skin color of all F_1 individuals was darker green than G35 and lighter green than Q51, but it inclined more towards dark green (Fig. 1a). Pigment content analysis showed that chlorophyll a and chlorophyll b contents were significantly lower in G35 than in Q51 (Fig. 1b). These results indicated that fruit skin color was determined by chlorophyll content.

Identification of a major QTL locus, CsFS1, on chromosome 3 by BSA-seq and GWAS

To rapidly identify loci for skin color in the F_2 population, two bulks consisting of 20 dark-green (SL-pool) and 20 light-green (QL-pool) progenies were sequenced on the Illumina platform. A total of 12.9 Gb of raw reads were generated, with an average depth of approximately 20.4×. The short reads were aligned to the cucumber reference genome [19], and 145,804 SNPs were identified between the dark-green and light-green parents. Based on the SNP-indices of the QL- and SL-pools, the Δ(SNP-index) of a genomic region from 36.62 Mb to 39.77 Mb on chromosome 3 was greater than the threshold value and close to 1.00 (Fig. 2a). This region may therefore harbor a major QTL for the fruit skin color trait in cucumber.

To independently confirm that this region was indeed related to fruit skin color, GWAS was performed on 289 cucumber accessions (average depth of 19.73× and 98.27% coverage of the cucumber reference genome) [19]. A total of 2,352,638 SNPs were identified using GATK software with default parameters [20]. To reduce the incidence of false-positive signals, a high-resolution variation map of 399,352 SNPs with minor allele frequency >5% and missing rate <0.2% was generated and used for genome-wide association analysis of fruit skin color with a unified mixed linear model that controlled for population structure and familial relatedness. A Manhattan plot for cucumber fruit skin color showed the strongest association signal (SNP_{fs}) on the distal arm of chromosome 3, overlapping with the genomic region identified by QTL-seq (Fig. 2b). This indicated that a major QTL controlling fruit skin color resided on the distal arm of chromosome 3, and it was named *CsFS1* (*Fruit skin 1*).

Fine mapping narrowed down CsFS1 to a 94-kb interval

To identify the candidate gene(s) in the *CsFS1* locus, classical QTL analysis was performed using 278 F_2 progenies. A total of 35 SNP markers were developed between 15.66 and 39.77 Mb on chromosome 3 and used for genotypic analysis of the F_2 segregating population (Table S3). QTL analysis using an MQM showed that the LOD peak from 64.85 to 69.05 cM was consistent with the physical distance from 39.0 to 39.77 Mb on chromosome 3 (Fig. 3a). In this interval, the highest LOD marker explained 35.6% of the phenotypic variation in the F_2 segregating population (Table S1). The genomic interval of *CsFS1* was further narrowed down to between two SNP markers (39,531,980 and 39,626,163 bp) using four recombinant individuals from the F_2 and BC_4F_2 populations (Fig. 3b). We therefore confirmed that the *CsFS1* locus lay within a 94-kb interval on chromosome 3.

Identification of a candidate gene related to fruit skin color

According to the cucumber genome database (http://www.icugi.org/), 12 of the 15 predicted protein-coding genes in the 94-kb interval have functional annotations (Table S2). qPCR experiments were performed to investigate the expression patterns of three possible candidate genes associated with fruit skin traits in NIL-1334 (light-green) and NIL-1325 (dark-green). In the pericarp, only the expression of *Csa3G912920* differed significantly between NIL-1334 and NIL-1325 ($P < 0.05$) (Fig. 3c, Fig. S2). The *Csa3G912920* gene encodes a plant GATA transcription factor and has a conserved zinc finger domain. A phylogenetic tree and sequence alignment showed that *Csa3G912920* homologs from melon (*MELO3C003335*), watermelon (*Cla97C09G175500*), and wax gourd (*Bhi05M000420*), highlighted in the gray-shadowed box, all encode GATA transcription factors (Fig. 4a and b). Secondary structural element analysis showed that the zinc finger domains include four β-folds and one α-spiral by looking up the literature (Fig. 4b). *Csa3G912920* was designated as a candidate gene for *CsFS1*.

Previous studies have shown that *Arabidopsis* GNC (GATA NITRATE-INDUCIBLE CARBON-METABOLISM-INVOLVED) and CGA1 (CYTOKININ-RESPONSIVE GATA1), members of the GATA transcription factor family, play a major role in the regulation of chlorophyll synthesis [21]. Under light, overexpression of GNC

promotes chloroplast development and the production of chlorophyll in roots [22]. We therefore inferred that *Csa3G912920* is the probable candidate gene for *CsFS1* and named it *CsGATA1*.

Discussion

In this study, we combined QTL-seq [23] of an F_2 segregating population with GWAS to identify a major QTL *CsFS1* for fruit skin color in cucumber. The main advantage of QTL-seq is that there is no need to develop DNA markers and marker genotyping. The SNP available between parental strains is such a marker, reducing cost and time. In addition, the use of SNP-index allows accurate assessment of the frequency of parental alleles. These advantages make QTL-seq an attractive method to quickly identify genomic regions containing major QTLs.

Fruit skin color is an essential agronomic trait in cucumber that affects exterior quality and consumer preferences. In this study, we detected the major QTL *CsFS1* on chromosome 3 between 39,531,980 and 39,626,163 bp. Previously, the *w* locus controlling the white fruit skin trait was also mapped to chromosome 3 (Liu et al. 2016), residing 281 kb upstream of the *CsFS1* locus. In the *w* locus, *Csa3G904140* (*APRR2*) harbors a single-nucleotide insertion that causes a frameshift mutation and a truncated protein in the white cucumber. Here, we found no sequence differences in *APRR2* between the two parents, G35 and Q51. Therefore, *CsFS1* is a novel QTL that controls green fruit skin in cucumber.

Through classical genetic mapping, *CsFS1* was narrowed to a 94-kb physical interval that contains 15 predicted protein-coding genes. The *Csa3G912920* gene encodes a GATA-type transcription factor, and its expression differed significantly between near isogenic lines with light- and dark-green fruit skins. Previous studies have shown that the GATA transcription factor families are highly conserved in *Arabidopsis*, rice, and other plants [24]. The GATA transcription factors are evolutionarily conserved transcriptional regulators that recognize promoter elements with a G-A-T-A core sequence [25]. The paralogous LLM-domain B-GATA transcription factors GNC and GNL contribute to chlorophyll biosynthesis and chloroplast formation in light-grown *Arabidopsis* seedlings [21, 26, 27]. Together, GNC and GNL control germination, greening, flowering time, and senescence downstream of auxin, cytokinin, gibberellin, and light signaling [28]. Studies have confirmed that some GATA genes are preferentially expressed in the leaf [29]. Leaves are the main organs for photosynthesis and light stress response in plants. High expression of a GATA transcription factor in leaves is consistent with its influence on chlorophyll synthesis. Therefore, it is reasonable to suggest that *Csa3G912920* is the candidate gene for fruit skin color in cucumber. Nonetheless, additional experiments are required to provide evidence for *Csa3G912920* gene function and robustly evaluate this hypothesis.

In conclusion, we identified a novel QTL, *CsFS1*, that controls green fruit skin color in cucumber and proposed a candidate gene, *Csa3G912920*, that may be responsible for the green color phenotype. Our results provide insight into the biological and molecular mechanisms of fruit skin color formation and can promote the development of attractive cucumber varieties with enhanced nutrients in the future.

Materials And Methods

Plant materials and phenotype evaluation

Two cucumber inbred lines, G35 (light-green skin color) and Q51 (dark-green skin color), were crossed to create F_1 progeny and then self-pollinated to generate an F_2 population. The F_1 progeny was backcrossed four times to the recurrent inbred parent G35 and then self-crossed to yield the BC_4F_2 generation. Chlorophyll a and chlorophyll b were extracted from fruit skins of G35, Q51, and F_1 progeny with ethyl alcohol and quantified by a spectrophotometric method. Two parental lines, together with the F_1 and F_2 generations, were used to describe and validate the inheritance pattern of skin color traits in immature fruit. Twenty F_2 individuals with extremely light-green skin color and 20 with extremely dark-green skin color were selected for BSA-seq. Two hundred seventy-eight individuals from the F_2 population were used for trait evaluation and QTL analysis. Fruit skin color in the F_2 population was independently evaluated by three persons. NIL-1334 (light-green skin) and NIL-1325 (dark-green skin) from the BC_4F_2 generation were used for gene expression analysis. Based on fruit skin color, 289 cucumber accessions were classified into eight categories (white, yellow-white, white-green, yellow-green, light-green, green, dark-green, and black-green) (Fig. S1) and used for GWAS analysis. All the plants were grown in plastic greenhouses under natural sunlight from spring 2016 to summer 2020 at the Tianjin Kernel Cucumber Research Institute.

Genomic DNA and total RNA extraction

Genomic DNA was extracted by the cetyltrimethylammonium bromide (CTAB) method [30] from fresh young leaves of P_1, P_2, and F_2 individuals and used for BSA-seq and QTL analyses.

Pericarp tissues were harvested from NIL-1334 and NIL-1325 at 0 days post-anthesis (DPA), 5 DPA, and 10 DPA. Each sample consisted of at least three fruits from different plants, and three replicate samples were used for gene expression analysis. Total RNA was extracted using the Quick RNA Isolation Kit (Huayueyang Biotechnology (Beijing) Co., Beijing, China) following the manufacturer's instructions. The concentration of total RNA was quantified using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Delaware, USA).

BSA-seq

Two DNA pools, the light-green pool (QL-pool) and dark-green pool (SL-pool), were created by mixing equal amounts of DNA from 20 individuals with light-green fruit skins and 20 individuals with dark-green fruit skins, respectively. Paired-end sequencing libraries (150-bp read length) with insert sizes of approximately 400 bp were prepared for sequencing on the Illumina NovaSeq 6000 platform. The short reads from the two pools were aligned to the reference genome of the 9930 line [19] using BWA software with default parameters [31]. SNP-calling was performed using SAMtools and BCFtools [31]. Low-quality SNPs with base quality value < 30, read depth < 2×, and mapping quality value < 30 were excluded to minimize false positives caused by repetitive genomic sequence or sequencing and alignment errors.

Two parameters, SNP-index and Δ(SNP-index) [23], were calculated to identify candidate regions for fruit skin color QTLs. SNP-index is the proportion of reads covering a given SNP that differ from the reference sequence. Thus, SNP-index = 0 if all short reads covering a given nucleotide position contain the reference SNP (9930 line), whereas SNP-index = 1 if all the short reads at that position contain the mutant SNP. Δ(SNP-index) is obtained by subtracting the SNP-index of the QL-pool from that of the SL-pool. The average SNP-index at a given genomic interval was calculated using a sliding window with a 1-Mb window size and a 10-kb increment. SNP-index graphs for the QL-pool and SL-pool, as well as the corresponding Δ(SNP-index), were plotted. The Δ(SNP-index) should not differ significantly from 0 in a genomic region with no major QTL [23]. We used a R script simulation to generate confidence intervals around the SNP-index under the null hypothesis of no QTL. First, we created two pools of progeny with a given number of individuals by random sampling. From each pool, a given number of alleles were sampled, corresponding to the read depth. Second, the SNP-index for each pool and the Δ(SNP-index) were calculated, and the process was iterated 10,000 times for each read depth to generate confidence intervals. Finally, these intervals were plotted for all genomic regions with variable read depths.

GWAS

Re-sequencing data from 289 cucumber accessions were obtained, with an average genome coverage of 98.27% and an average sequencing depth of 19.728×. We obtained 2,352,638 SNPs, and 399,352 high-quality SNPs were retained, with a deletion rate of less than 0.2. The association between fruit skin color and each SNP was tested using a unified mixed model [32, 33] that includes principal components [34] as a fixed effect to account for the population structure and kinship matrix [35] and to explain familial relatedness. Using the Bayesian information criterion, a backward elimination procedure was implemented to determine the optimal number of principal components to include in the mixed model [36]. The false discovery rate was controlled at 5% using the Benjamini and Hochberg procedure [37]. A likelihood ratio-based r^2 statistic was used to assess the goodness-of-fit of each SNP [38]. All analyses were performed using the Genome Association and Prediction Integrated Tool (GAPIT) package [39].

Marker development and QTL analysis

The SNPs were filtered from the re-sequencing data of the two parents, G35 and Q51. The sequence context of the candidate SNPs was examined in the 9930 reference genome using BLAST alignment to obtain longer sequences for marker development. In total, 35 kompetitive allele specific PCR (KASP) SNP markers on chromosome 3 were developed using the BSA-seq and GWAS data and created using Primer 5.0 (PREMIER Biosoft International, USA) (Table S3). The genotypes of the F_2 population were analyzed using an Infinite M1000 microplate reader (Tecan, Switzerland) and the online tool “snpdecoder” (http://www.snpway.com/snpdecoder/). Linkage analysis was performed using JoinMap 4.0 [40], and QTL analysis was performed in MapQTL6.0 using the multiple QTL model (MQM mapping) procedure [41](Van Ooijen, 2009).

Quantitative real-time PCR (qRT-PCR)

Single-stranded cDNA was synthesized using the PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa Bio Inc., Dalian, China) following the manufacturer's instructions. qRT-PCR was performed in a 10-μl reaction volume consisting of 5 μl TB Green Premix Ex Taq (Tli RNaseH Plus) (TaKaRa), 0.25 μl ROX Reference Dye (50×), 0.25 μl each of forward and reverse primers (10 μM), 1 μl cDNA templates, and 3.25 μl purified water. Thermal cycling began with an initial step at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 34 s, and it was performed on the QuantStudio Flex 6 Real-Time PCR System (Applied Biosystems, California, USA). All samples were performed in triplicate, and *CsACTIN* (*Csa2G018090*) was used as the internal reference gene. Relative expression values were determined using the comparative Ct method ($2^{-\Delta\Delta Ct}$). Primers used for qRT-PCR are listed in Table S4.

Phylogenetic analysis

CsGATA1 and its homologous amino acid sequences were retrieved from public databases: SolGenomics (https://solgenomics.net/) and the Cucurbit Genomics Database (http://cucurbitgenomics.org). Known GATA transcription factors from rice, maize, and *Arabidopsis* were added to the analysis. Sequence alignments and a neighbor-joining tree with 1000 bootstrap replicates were constructed in MEGA X [42].

Declarations

Ethics approval and consent to participate

With the permission to collect, all materials of *Cucumis sativus* accessions were identified and collected from China, and now deposited at Tianjin Kernel Cucumber Research Institute, China. The study complied with relevant institutional, national, and international guidelines and legislation. This research did not involve any human subjects, human material, or human data. *Cucumis sativus* in current research did not belong to the endangered or protected species.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The raw Illumina sequence reads have been deposited into the National Genomics Data Center (https://bigd.big.ac.cn/) under accession number CRA004282.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding

This work was supported by the National Key Research and Development Program of China (2016YFD0101900) and the Tianjin Natural Science Foundation Project (20JCYBJC00720).

Authors' contribution

YHL and HYH designed the research; XLG, LDZ, and AMW performed the experiments and analyzed the data; QQY and HZW analyzed the data; WLK, JWL, SLD, RHY, and HYH conducted the field trials; QQY, XLG, SLD, RHY and TL wrote the manuscript, and all authors read and approved the manuscript.

Acknowledgments

The authors would like to thank TopEdit (www.topeditsci.com) for linguistic assistance during preparation of this manuscript.

References

1. Freilich S, Lev S, Gonda I, Reuveni E, Portnoy V, Oren E, Lohse M, Galpaz N, Bar E, Tzuri G *et al*: Systems approach for exploring the intricate associations between sweetness, color and aroma in melon fruits. *BMC Plant Biol* 2015, 15:71.
2. Feder A, Burger J, Gao S, Lewinsohn E, Katzir N, Schaffer AA, Meir A, Davidovich-Rikanati R, Portnoy V, Gal-On A *et al*: A Kelch Domain-Containing F-Box Coding Gene Negatively Regulates Flavonoid Accumulation in Muskmelon. *Plant Physiol* 2015, 169(3):1714–1726.
3. Zhao G, Lian Q, Zhang Z, Fu Q, He Y, Ma S, Ruggieri V, Monforte AJ, Wang P, Julca I *et al*: A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat Genet* 2019, 51(11):1607–1615.
4. Li B, Lu X, Dou J, Aslam A, Gao L, Zhao S, He N, Liu W: 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* 2018, 19(10).
5. Dou J, Lu X, Ali A, Zhao S, Zhang L, He N, Liu W: Genetic mapping reveals a marker for yellow skin in watermelon (Citrullus lanatus L.). *PLoS One* 2018, 13(9):e0200617.
6. Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, Yativ M, Domínguez E, Wang Z, De Vos RCH, Jetter R *et al*: Fruit-Surface Flavonoid Accumulation in Tomato Is Controlled by a SlMYB12-Regulated Transcriptional Network. *PLOS Genetics* 2009, 5(12):e1000777.

7. Ballester A-R, Molthoff J, de Vos R, Hekkert BtL, Orzaez D, Fernández-Moreno J-P, Tripodi P, Grandillo S, Martin C, Heldens J *et al*: Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. *Plant Physiology* 2010, 152(1):71–84.
8. Hurtado-Hernandez H, Smith PG: Inheritance of mature fruit color in Capsicum annuum L. *Journal of Heredity* 1985, 76(3):211–213.
9. Liu H, Meng H, Pan Y, Liang X, Jiao J, Li Y, Chen S, Cheng Z: 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* 2015, 128(12):2375–2385.
10. Anne, Cortleven, Thomas, Schmülling: Regulation of chloroplast development and function by cytokinin. *Journal of experimental botany* 2015.
11. Jiao J, Liu H, Liu J, Cui M, Xu J, Meng H, Li Y, Chen S, Cheng Z: Identification and functional characterization of APRR2 controlling green immature fruit color in cucumber (Cucumis sativus L.). *Plant Growth Regulation* 2017, 83(2):233–243.
12. Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW: ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci U S A* 2003, 100(7):4328–4333.
13. Zhou Q, Wang S, Hu B, Chen H, Zhang Z, Huang S: An ACCUMULATION AND REPLICATION OF CHLOROPLASTS 5 gene mutation confers light green peel in cucumber. *J Integr Plant Biol* 2015, 57(11):936–942.
14. Bollivar D, Braumann I, Berendt K, Gough SP, Hansson M: The Ycf54 protein is part of the membrane component of Mg-protoporphyrin IX monomethyl ester cyclase from barley (Hordeum vulgare L.). *Febs j* 2014, 281(10):2377–2386.
15. Lun Y, Wang X, Zhang C, Yang L, Gao D, Chen H, Huang S: A CsYcf54 variant conferring light green coloration in cucumber. *Euphytica* 2015, 208(3):509–517.
16. Hao N, Du Y, Li H, Wang C, Wang C, Gong S, Zhou S, Wu T: CsMYB36 is involved in the formation of yellow green peel in cucumber (Cucumis sativus L.). *Theoretical and Applied Genetics* 2018, 131(8):1659–1669.
17. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al*: Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* 2010, 465(7298):627–631.
18. Michelmore RW, Paran I, Kesseli RV: 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* 1991, 88(21):9828–9832.
19. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al*: The genome of the cucumber, Cucumis sativus L. *Nat Genet* 2009, 41(12):1275–1281.
20. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-

generation DNA sequencing data. *Genome Res* 2010, 20(9):1297–1303.

21. Chiang YH, Zubo YO, Tapken W, Kim HJ, Lavanway AM, Howard L, Pilon M, Kieber JJ, Schaller GE: Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiol* 2012, 160(1):332–348.
22. Richter R, Bastakis E, Schwechheimer C: Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. *Plant Physiol* 2013, 162(4):1992–2004.
23. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S *et al*: QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 2013, 74(1):174–183.
24. Reyes JC, Muro-Pastor MI, Florencio FJ: The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiol* 2004, 134(4):1718–1732.
25. Behringer C, Schwechheimer C: B-GATA transcription factors - insights into their structure, regulation, and role in plant development. *Front Plant Sci* 2015, 6:90.
26. Bastakis E, Hedtke B, Klermund C, Grimm B, Schwechheimer C: LLM-Domain B-GATA Transcription Factors Play Multifaceted Roles in Controlling Greening in Arabidopsis. *Plant Cell* 2018, 30(3):582–599.
27. Bi YM, Zhang Y, Signorelli T, Zhao R, Zhu T, Rothstein S: Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. *Plant J* 2005, 44(4):680–692.
28. Ranftl QL, Bastakis E, Klermund C, Schwechheimer C: LLM-Domain Containing B-GATA Factors Control Different Aspects of Cytokinin-Regulated Development in Arabidopsis thaliana. *Plant Physiol* 2016, 170(4):2295–2311.
29. Ao T, Liao X, Xu W, Liu A: Identification and characterization of GATA gene family in castor bean (Ricinus communis). *Plant Diversity and Resources* 2015, 37(4):453–462.
30. Murray MG, Thompson WF: Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 1980, 8(19):4321–4325.
31. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25(14):1754–1760.
32. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 2006, 38(2):203–208.
33. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM *et al*: Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 2010, 42(4):355–360.
34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006, 38(8):904–909.

35. Loiselle BA, Sork VL, Nason J, Graham C: Spatial genetic structure of a tropical understory shrub, PSYCHOTRIA OFFICINALIS (RuBIACEAE). *American Journal of Botany* 1995, 82(11):1420–1425.
36. Schwarz G: Estimating the Dimension of a Model. *Ann Statist* 1978, 6(2):461–464.
37. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995, 57(1):289–300.
38. Sun G, Zhu C, Kramer MH, Yang SS, Song W, Piepho HP, Yu J: Variation explained in mixed-model association mapping. *Heredity (Edinb)* 2010, 105(4):333–340.
39. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: GAPIT: genome association and prediction integrated tool. *Bioinformatics* 2012, 28(18):2397–2399.
40. Ooigen V: Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res (Camb)* 2011, 93(5):343–349.
41. van Der Schaar W, Alonso-Blanco C, Léon-Kloosterziel KM, Jansen RC, van Ooijen JW, Koornneef M: QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. *Heredity (Edinb)* 1997, 79 (Pt 2):190–200.
42. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016, 33(7):1870–1874.

Figures

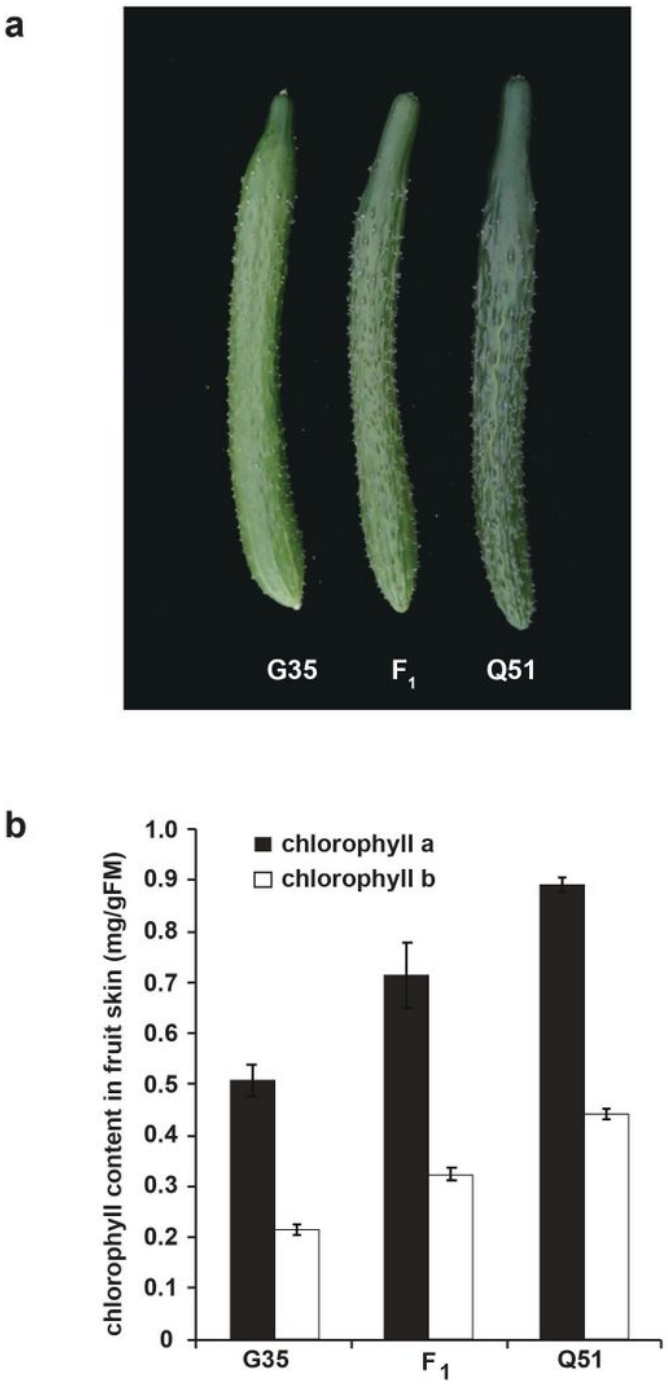

Figure 1

The fruit skin color traits of two parents and their F_1 hybrid.

a G35 (P1, left), an F_1 hybrid of G35 × Q51 (middle), and Q51 (P2, right). Photos of cucumber fruit were taken 10 days post-anthesis (DPA). **b** The content of chlorophyll a and b in two parents (G35 and Q51) and their F_1 hybrid.

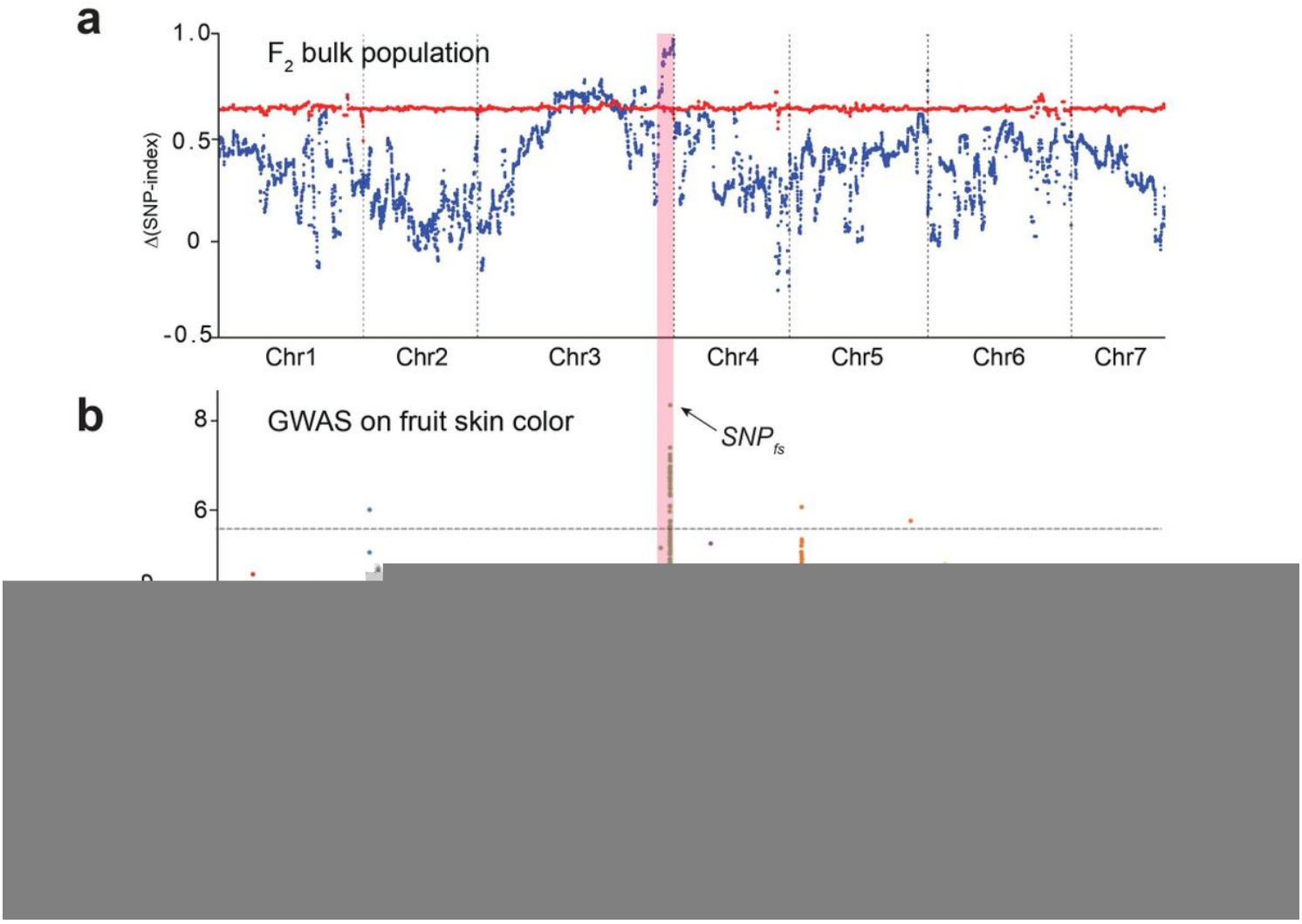

Figure 2

Identification of overlapping intervals identified by BSA-seq and GWAS for fruit skin color in cucumber.

a Δ(SNP-index) plot with statistical confidence intervals under the null hypothesis of no QTL (red, $P < 0.01$). The candidate QTL (*CsFS1*) location was identified between 36.62 and 39.77 Mb on chromosome 3. **b** GWAS analysis (Manhattan plots) showed a significant peak (SNP_{fs}) above the threshold on chromosome 3 within the region previously identified in the QTL-seq analysis.

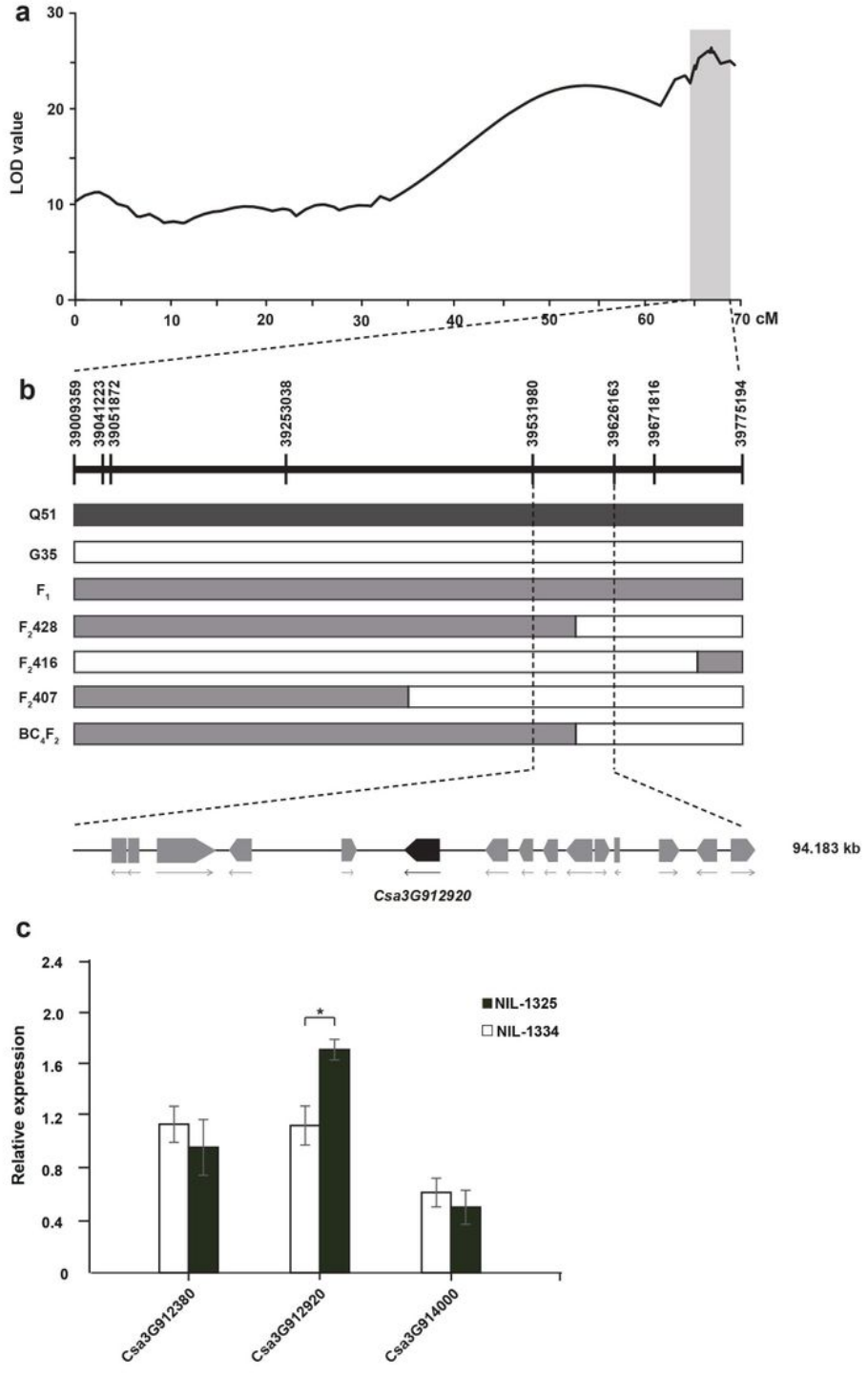

Figure 3

Fine mapping of *CsFS1* on chromosome 3.

a LOD (log 10 of the odds ratio) plots of linkage analysis based on SNP markers indicate the most likely position of *CsFS1* between markers SNP39009359 and SNP39775194 on chromosome 3. **b** Mapping of the *CsFS1* region using three recombinants with extremely light-green fruit skin color identified from 278

plants in the F_2 and BC_4F_2 populations. *CsFS1* was placed within a 94-kb interval containing 15 candidate genes between the markers SNP39531980 and SNP39626163. **c** Relative expression of three candidate genes in the fruit pericarp of the light-green near isogenic line NIL-1334 and the dark-green near isogenic line NIL-13250 at 0 days post-anthesis (DPA). The relative expression is shown as the mean ± standard deviation, and statistical significance was determined using Student's *t*-tests (*$P < 0.05$).

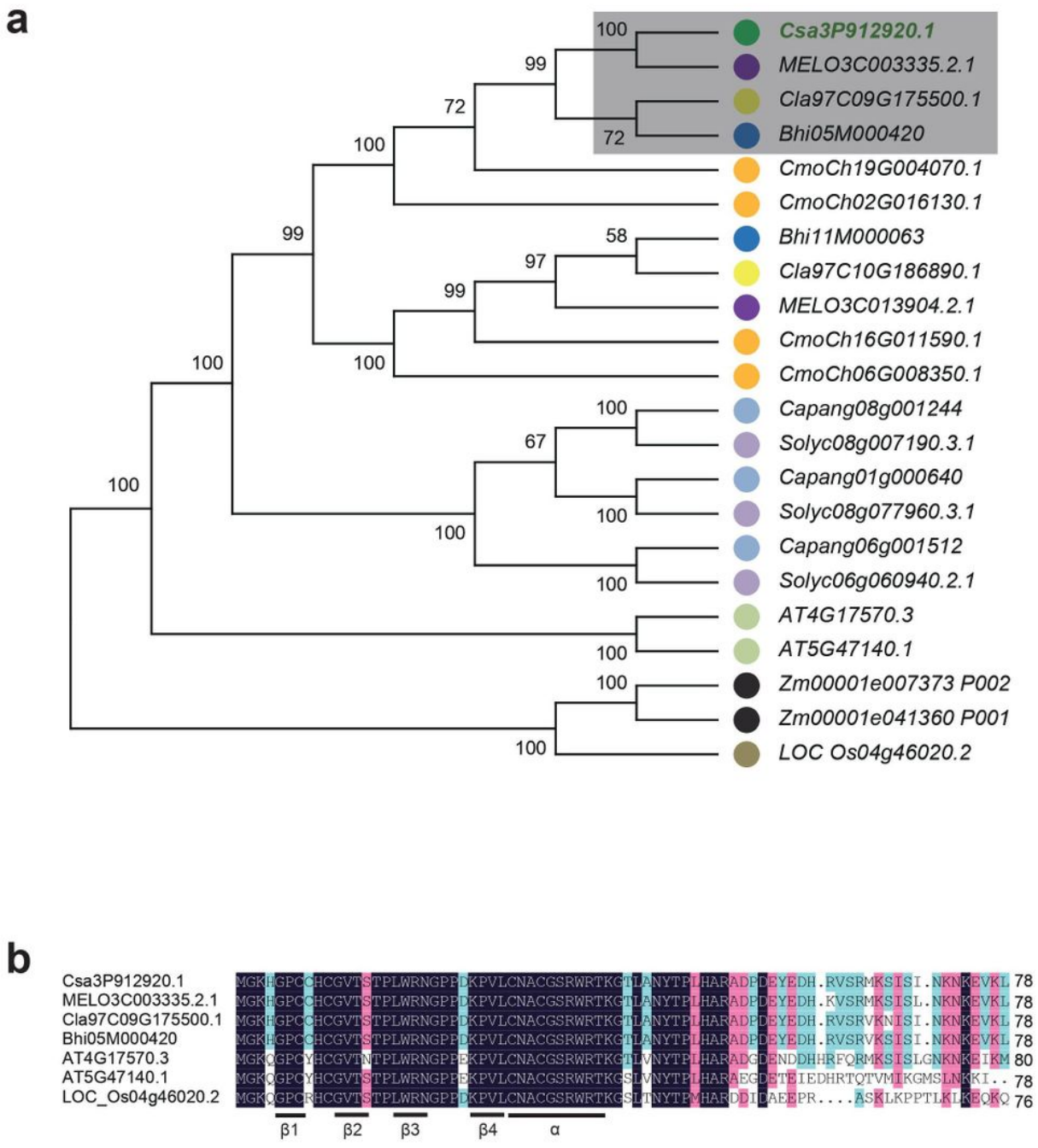

Figure 4

Phylogenetic tree and structure identity of *Csa3G912920* and its homologs in different species.

a Phylogenetic tree of Csa3G912920 and its homologs in *Arabidopsis*, rice, maize, melon, watermelon, pumpkin, wax gourd, tomato, and pepper. The closest homologs of Csa3G912920 are indicated in a gray-shadowed box and include those from melon (MELO3C003335.2.1), watermelon (Cla97C09G175500.1), and wax gourd (Bhi05M000420). **b** Alignment of Csa3G912920, MELO3C003335.2.1, Cla97C09G175500.1, Bhi05M000420, AT4G17570.3, AT4G47140.1, and LOC_Os04g46020.2 protein sequences. Amino acid residues with at least 70.51% identity or similarity between these homologs are shaded black or red or blue, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles.docx
- FigS1.pdf
- FigS2.pdf
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx