Candidate gene analysis of watermelon stripe pattern locus ClSP ongoing recombination suppression

Zhen Yue
Northwest A&F University: Northwest Agriculture and Forestry University
Rongxue Ma
Northwest A&F University: Northwest Agriculture and Forestry University
Denghu Cheng
Northwest A&F University: Northwest Agriculture and Forestry University
Xing Yan
Northwest A&F University: Northwest Agriculture and Forestry University
Yaping He
Northwest A&F University: Northwest Agriculture and Forestry University
Chunxia Wang
Northwest A&F University: Northwest Agriculture and Forestry University
Xiaona Pan
Northwest A&F University: Northwest Agriculture and Forestry University
Lijuan Yin
Northwest A&F University: Northwest Agriculture and Forestry University
Xian Zhang
Northwest A&F University: Northwest Agriculture and Forestry University
Chunhua Wei (xjwend020405@163.com)
Northwest Agriculture and Forestry University https://orcid.org/0000-0002-7110-4195

Research Article

Keywords: CISP, watermelon stripe pattern, transcriptome analysis

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-219409/v1

Abstract

Stripe pattern is an important commodity trait in watermelon, displaying diverse types. In this study, two segregating populations were generated for genetic mapping the single dominant locus *ClSP*, which was finally delimited to a 611.78 Kb interval with suppression of recombination. According to polymorphism sites detected among genotypes, four discrete haploblocks were characterized in this target region. Based on reference genomes, 81 predicted genes were annotated in the *ClSP* interval, including seven transcription factors namely as candidate No1-No7. Meanwhile, the ortholog gene of cucumber *ist* responsible for the irregular stripes was considered as candidate No8. Strikingly, gene structures of No1-No5 completely varied from their reference descriptions and subsequently re-annotated. Notably, the adjacent distribution candidates No2 and No3, as well as No4 and No5, were confirmed to derive from longer transcripts designated as No2_3 and No4_5, respectively. Sequence analysis demonstrated the third polymorphism in CDS of re-annotated No4_5 resulting in truncated proteins in non-stripe plants. Furthermore, only No4_5 was down-regulated in non-stripes relative to stripes contrast to other candidates. Transcriptome analysis identified 356 DEGs between striped and non-striped peels, with genes involved in photosynthesis and chloroplast development down-regulated in non-stripes but calcium ion binding related genes up-regulated. Additionally, 38 DEGs were annotated as transcription factors, with the majority up-regulated in non-stripes, such as ERFs and WRKYs. This study not only contributes to a better understanding of the molecular mechanisms underlying watermelon stripe development, but also provides new insight into the genomic structure of *ClSP* locus and valuable candidates.

Introduction

Watermelon (*Citrullus lanatus* L., 2n = 2x =22) belonging to Cucurbitaceae family, is cultivated world widely and consumed as one of the most favorite fresh fruits globally. Due to its important economic and horticultural performance, watermelon breeders have contributed to improve the fruit quality and morphological characteristics. As an important agronomic appearance and commodity trait, stripe pattern (morphological trait) can directly affect the evaluation standard of consumers in the market (Lou and Wehner 2016). For example, the American origin cultivar Crimson Sweet as well as the similar types (light green stripe with broad width and diffused margin) are very popular in USA, Brazil and Europe, while the Jubilee-type varieties (dark green stripes with medium width and sharp margin) are widely cultivated in China and South Korea (Gama et al. 2015; Kim et al. 2015; Park and Cho 2012). Since the 1930s, genetic studies have been performed to analyze the inheritance of watermelon rind traits, such as stripe patterns, rind colors, and fruit shapes (Poole 1944; Rhodes and Zhang 1995; Weetman 1937; Wehner and Pitrat 2008). According to the morphological descriptions of watermelon, rind types could be further divided into background rind color (light to dark green, yellow, or gray) and foreground stripe pattern (solid to stripe), the latter of which are diverse and further characterized in terms of the margin of stripes (well-defined, medium, and diffuse), the width of stripes (very narrow to very broad), the intensity of stripe coloration (unicolored, bicolored, marbled, or only vein), the conspicuousness of stripes (inconspicuous to very strong) (Gama et al. 2015; Kim et al. 2015). In the late 1930s and early 1940s, complex models of

inheritance of rind pattern in watermelon have been raised. Researchers Weetman and Poole hypothesized that three alleles (*G*, g^s, and *g*) at the *g* locus determine the inheritance of rind colors and stripes (Poole 1944; Weetman 1937). Gene g^s is proposed to control the presence of rind stripes, which is recessive to the allele *G* (dark green rind color) but dominant to the light green rind color gene *g*. Recently, a more complex model was raised to explain the inheritance of fruit rind colors and stripe patterns, in which contains five alleles at the *g* locus, *G* (dark green rind color), g^W (wide stripe), g^M (medium stripe), g^N (narrow stripe), and *g* (light green rind color), with the dominance $G > g^W > g^M > g^N > g$ (Lou and Wehner 2016). Compared to narrow stripes, the wide stripe pattern is proposed to be dominant and regulated by a single gene (Lou and Wehner 2016). Moreover, the blurred stripe pattern is dominant over the clear stripe margin phenotype and controlled by a single gene *Csm* (Lou and Wehner 2016), while the qualitative character of well-defined stripes is dominant to the diffused stripe pattern (Gama et al. 2015). Regarding of other stripe patterns in watermelon, the intermittent rind phenotype of 'Navajo Sweet' showing narrow dark green stripes at the stem end of the fruit but being irregular around the fruit equator and nearly absent at the blossom end of the fruit, is regulated by a single recessive gene *ins* (Gusmini and Wehner 2006); the penciled (p) phenotype visualized as very narrow stripes on a light background rind of 'Japan 6' is recessive to the netted rind pattern of 'China 23' (Weetman 1937).

Compared to one-locus multiple-allele model, three genetically independent loci, *S* (foreground stripe pattern), *D* (depth of rind color), and *Dgo* (background rind color), were proposed to determine the rind colors and stripe patterns in watermelon, which were independently located on chromosomes 6, 8 and 4, respectively (HeeBum et al. 2015; Park et al. 2016). Using the GWAS method, the *Cla97C06G126710* gene encoding a WD40-repeat protein was identified as the best likely candidate gene, which is predicted to control the presence of stripe rind of watermelon (Guo et al. 2019). On the contrary, a single gene locating on chromosome 8 is hypothesized to be responsible for the dominant wide stripe pattern in watermelon (Zhang et al. 2018). As an important trait in watermelon breeding, linkage markers relative to stripe rind pattern have been explored for marker assisted breeding. For instance, a SCAR marker wsbin6-11 at the physical location of 23.32 Mb on chromosome 6 was developed to be tightly linked with JT stripe pattern (dark green with medium width and sharp margin) (Kim et al. 2015). Two microsatellite loci MCPI_05 and MCPI_16 with linkage to the stripe pattern were also located on the chromosome 6 of watermelon genome (Gama et al. 2015). Representing as a secondary skin color, the diverse rind stripes were also widely investigated in other cucurbit crops. For example, the striped rind dominant gene *st3* has been fine mapped in a 172.8 Kb region on chromosome 4 in melon (Liu et al. 2019), and gene *Cmsp-1* controlling the dominant spotted trait has been located in a 3.94 Mb region on the end of chromosome 2 (Lv et al. 2018). In cucumber, gene *Csa1G005490* encoding a polygalacturonase-1 noncatalytic subunit beta protein (PG1β) was identified as the possible candidate gene for the irregularly striped rind pattern in mutant *ist* (Song et al. 2019). The stripe development in Cucurbit species is also a complex process, which results from the action of multiple genes, such as genes *l-1*, $l\text{-}1^{Bst}$, $l\text{-}1^{St}$, $l\text{-}1^{iSt}$, and *l-2* (Grumet and Colle 2016; Paris 2003, 2009). In other horticultural species, the stripe rind patterns are also investigated. For instance, the red and green stripes of apple peels are due to the differences of anthocyanin accumulation, which are associated with transcript abundances of the *MYB10* transcription factor (Telias

et al. 2011), similar to that in pear (Qian et al. 2014). In tomato, green stripe phenotype is caused by the high degrees of methylation of the *TAGL1* promoter, which encodes a MADS-box transcription factor (Liu et al. 2020).

In this study, we recruited two segregating populations to map the causal locus determining the Jubilee-type similar stripe pattern (dark green stripes with medium width and sharp margin). Inheritance analyses revealed that this rind appearance was genetically controlled by a dominant gene *ClSP* (*Citrullus lanatusStripe Pattern*), which was finally delimited in a 611.78 Kb region on chromosome 6 according to the reference genome 797103 (V2). Interestingly, at least four discrete haplotype blocks were characterized in the mapping interval basing our RNA-seq and DNA-seq data, which contained seven transcription factor candidates (No1-7) and the ortholog (candidate No8) to cucumber gene *ist* responsible for the irregularly striped rind. Furthermore, gene structure, sequence polymorphism, and expression accumulation of these candidates were analyzed. Additionally, enrichment analyses of DEGs identified through transcriptome analysis provide new insights into the molecular mechanisms underlying watermelon stripe development.

Materials And Methods

Plant materials and phenotypic characterization

In our previous study (Wei et al. 2017), a larger-sized $F_{3:4}$ population derived from commercial watermelon hybrid cultivar 'Lingxiu' was used to fine map the lobed leaf gene *ClLL1*, of which individuals also exhibited two distinct rind phenotypes (Fig. 1). The majority displayed a distinct rind pattern, dark green stripes with clear margins on a standard green background, while the rest exhibited standard green rind with netted reticulations on the whole fruit surface. These two distinguishable stripe patterns were defined as stripe and non-stripe phenotypes in this study, referring to the highly similar morphological rind features described recently (HeeBum et al. 2015). Apart from this $F_{3:4}$ population, another F_2 segregating population derived from a cross between inbred lines M08 and N21 was also used for stripe pattern inheritance analysis and linkage mapping of stripe locus *ClSP*. The round fruit of parent line M08 has dark green stripes with clear margins on green background rind, while the parent N21 has elongate fruit with inconspicuous reticulations, which have also been recruited as parent lines in dwarfism phenotype analysis (Wei et al. 2019). All these materials were grown in greenhouses of Northwest A&F University, Yangling, China. The phenotypes of stripe pattern were visually distinguished at 20-30 days after pollination (DAP), and subsequently recorded as stripe or non-stripe pattern. The deviation from the expected 3:1 segregation ratio in the $F_{3:4}$ (n = 1286) or F_2 population (n = 358) was tested using the chi-square test.

Bulked segregant analysis and RNA-seq

Using the combing approach of BSA and RNA-seq (BSR-seq), 34 striped and non-striped pattern plants were randomly selected, and approximately one square centimeter of rind at the stripes or netted reticulations were sampled to extract total RNA. After removing the contaminating genomic DNA, equivalent total RNA of 34 striped and non-striped samples were independently pooled, constituting the stripe and non-stripe bulks, respectively. RNA sequencing generating 150 bp paired-end reads was performed on Illumina HiSeq™ X platform. Clean RNA-seq data were mapped onto the watermelon reference genome (97103, V2) with the software Bowtie2, and SNPs calling was performed with software SAMtools (Guo et al. 2019; Langmead and Salzberg 2012; Li et al. 2009). Subsequently, the frequency for each SNP was calculated for both pools via in-house Perl scripts following the published workflow (Su et al. 2019; Yan et al. 2019), which was in turn used to calculate the △(SNP-index) by subtracting the SNP-index of striped pool from that of non-striped pool. After plotting the average values of △ SNP-index on chromosomes, the region with a high △(SNP-index) was primarily considered as the potential target interval harboring the *ClSP* locus.

To explore the underlying mechanisms of stripes formation, the peel samples from stripes and non-stripes rind were precisely isolated for RNA extraction. Total RNA from two biological replicates were designated as Stripe_1, _2 and Non-stirpe_1, _2 respectively, and subsequently sent for sequencing on a Novaseq6000 platform (Novogene, Guangzhou, China). After discarding the low quality reads, clean data were mapped onto the watermelon reference genome 97103 (Guo et al. 2019). Genes with expression changes more than two fold (p-adjust < 0.05) were considered as differentially expressed genes (DEGs). The subsequent GO and KEGG functional enrichment analyses of DEGs were conducted using tools on a free online platform OmicShare (www.omicshare.com). In addition, the genome re-sequencing data of two parental lines M08 and N21 had also been used for marker development and sequence polymorphism analysis, and the detailed bioinformatics analysis has been described in our published research (Wei et al. 2019).

Genetic mapping of *ClSP* locus

According to the prominent peak drawn from BSR-seq analysis, five polymorphic markers were designed in the potential region and were used to screen the $F_{3:4}$ segregating population (n = 119). After confirming the presence of locus *ClSP*, seven new markers in primary interval were developed to genotype a large $F_{3:4}$ population with 1167 individuals. Finally, the locus *ClSP* was delimited in a 794.03 region between markers W07061 and W07134, with 38 and 26 recombinants respectively, inferring a severe suppression of recombination. To further narrow down the mapping interval and confirm the presence of recombination suppression, another segregating F_2 population (n = 1594) was generated from a cross between M08 (stripe pattern) and N21 (non-stripe pattern). After polymorphic examination, four markers from the $F_{3:4}$ population also exhibited efficiency between parent lines M08 and N21. Combined with three new markers, the locus *ClSP* was mapped in a 611.78 Kb region between markers W11041 and W07093, with 14 and 2 recombinants respectively. Totally, five genetic markers were co-segregated with

the related phenotypes, which can be further used in marker assisted breeding for watermelon stripe patterns. All the primer sequences used in this mapping strategy were listed in Table S1.

Gene annotation and hyplotype analysis of *ClSP* locus

The annotated genes in the mapping interval were retrieved according the watermelon reference genome 97103 (V2) (Guo et al. 2019). To explore all the possible candidate genes in the target region, all polymorphic markers were mapped on the other versions of watermelon reference genome using Blast program, i.e., genomes 97103 (V1) and Charleston gray (V1) (Guo et al. 2013; Wu et al. 2019). The redundant genes from these three version reference genomes were removed based on the orthologous relationship (http://cucurbitgenomics.org/). To analyze the hyplotype structure of the *ClSP* locus among the five genotypes (reference 97103, stripe and non-stripe pools, M08 and N21) used in this study, genomic polymorphisms were visually compared using the BSR-seq and DNA-seq data. Hyplotype blocks were subsequently distinguished according the sequence polymorphic variations among genotypes.

Gene structure and sequence analysis of candidate genes

Gene structures of all candidates were firstly validated using the BSR-seq data of stripe and non-stripe pools. Then, to verify the predicted gene structures, gene specific primers were designed to amplify the nearly full-length coding sequences. Give that multiple transcripts were identified for some candidate genes, the longest one was used for further sequence polymorphism analysis. The primer information for gene cloning was also listed in Table S1, and their relative positions on candidates were presented in Fig. S3.

To analyze polymorphic variations of candidate genes among genotypes, the re-sequencing data of two parental lines N21 and M08, as well as the BSR-seq data of two pools, were mapped onto the high quality reference genome 97103 (V2) (Guo et al. 2019). Then, the mapped reads on each candidate were visually compared using IGV software (Thorvaldsdottir et al. 2013). Additionally, the polymorphic sites in coding sequence of candidates between stripe and non-stripe pools were also confirmed by gene cloning assays mentioned above.

RNA isolation and qRT-PCR

Total RNA were extracted from the precisely sampled tissues from stripped and non-stripped peels using the RNA Simple Total RNA Kit (TIANGEN, China), and the first strand cDNA were subsequently synthesized via the FastKing RT Kit with gDNase (TIANGEN, China), following the manufacture's protocol. The qRT-PCR amplification was performed on a StepOnePlus Real-Time PCR system (Applied Biosystems, Foster, USA). Using housekeeping gene *ClACT* (*Cla007792*) as internal reference, the relative expression level for each selected gene was calculated using the $2^{-\Delta\Delta Ct}$ method, as described in our previous study (Wei et al. 2019).

In order to examine the expression profile of candidates during stripe formation, the gene-specific primers were designed according to their coding sequences cloned from the stripe pool. Meanwhile, to validate the reliability of the four RNA-seq data, twenty DEGs were selected, and their specific primers were designed according to their reference coding sequences. All the primers used in qRT-PCR experiments were also listed in Table S1.

Results

Inheritance analysis of watermelon stripe patterns

In our previous work, the lobed leaf gene *ClLL1* has been fine mapped of using a $F_{3:4}$ population (Wei et al. 2017), in which the rind stripe phenotype is also segregated (Fig. 1). During the early developmental stages of ovary, penciled stripes were regularly distributed over the pericarp covering by trichome. However, at approximately 6 days after pollination (DAP, D6 in Fig. 1), the stripe pattern could be visually characterized and distinctly distinguished at 8 DAP (D8), which was subsequently defined as standard green rind with stripes or without stripes at mature fruit period (D23) according to the descriptions of similar rind phenotypes published recently (HeeBum et al. 2015; Park et al. 2016). In a 1286 $F_{3:4}$ segregated population, there were 987 individuals with stripes and 299 without stripes (Table 1), fitting a 3:1 Mendelian ratio ($c^2 = 2.01$, $p = 0.15$), which inferred that the stripe phenotype is controlled by a single dominant gene.

To confirm this assumption, an inbred line M08 with standard green rind color and black stripes was used to cross with a dark green rind color and non-stripes line N21 to generate a new F_2 population (Fig. S1). The fruit phenotypes of F_2 offspring varied dramatically, including fruit shape, background color of rind, and stripe patterns. However, the latter could be clearly classified as stripes and non-stripes at mature fruit period (Fig. S1). Phenotypic data revealed that 358 individuals from F_2 population contained 281 stripe and 77 non-stripe plants (Table 1), also consistent with the Mendelian ratio of 3:1 ($c^2 = 2.15$, $p = 0.13$). Taken together, all these data suggested that the stripe phenotype in watermelon is controlled by a single dominant gene and designated as *ClSP* (*Citrullus lanatusStripe Pattern*) hereinafter.

Potential location of *ClSP* locus by BSR-seq analysis

As shown by morphological observations (Fig. 1), the stripe patterns could be visually distinguished at 8 DAP (D8) in the $F_{3:4}$ population. To preliminarily located locus *ClSP* on chromosome, the method combining BSA and RNA-seq was used in this study. We separately selected 34 stripe and non-stripe pattern young fruits 8 days after pollination from the $F_{3:4}$ population. Then, total RNA were independently extracted using approximately one square centimeter of fruit peel (Fig. 1), and subsequently bulked as stripe pool and non-stripe pool for sequencing. A total of 43.9 (6.59 Gb) and 51.1 million clean reads (7.67 Gb) were generated from stripe and non-stripe bulks, with Q30 values more than 91.00% (Table 2). After mapping to watermelon reference genome 97013 (V2), single nucleotide polymorphisms (SNPs)

were called, and subsequently the allele frequencies (index) for each SNP were calculated for each bulk. According to the average values of delta SNP index plotting on chromosomes, we obtained a prominent peak at the end part of chromosome 6 (Fig. 2a), suggesting the possible presence of the *CISP* in this corresponding region.

Genetic mapping of *CISP* locus

To verify the potential location of *CISP* locus, three polymorphic markers (W07302, W07094, and W07096) were developed to screen a small $F_{3:4}$ segregating population with 119 individuals. Linkage analysis indicated that the *CISP* locus was delimited in a 1.48 Mb region between markers W07302 and W07094 with thirteen and two recombinants, respectively (Fig. 2b). Another two new markers FD01071 and W07092 were designed and confirmed to be co-segregate with the object phenotype. To further narrow down this target region, a larger $F_{3:4}$ segregating population consisting of 1167 individuals was utilized to genotype with the primary flanking markers W07302 and W07094, and 132 and 27 new recombinants were identified, respectively. Combined with the recombinants obtained in the initial small population, a total of 172 were used for further analysis. Consequently, seven new markers were designed in the primary region, and were used to genotype the 172 recombinants. Finally, gene *CISP* was finally delimited in a 794.03 Kb region between markers W07061 and W07134, with 38 and 26 recombinants respectively (Fig. 2c), inferring suppression of recombination present in this interval. Due to lack of effective recombinants, further narrowing down of this mapping region was unfeasible and the five markers (W07132, W07163, W07062, FD01071, and W07092) were co-segregated with stripe patterns, which can be used for marker selection breeding program.

To further delineate this target region, a new F_2 segregating population was generated by crossing inbred lines M08 (stripe pattern) with N21 (non-stripe pattern). The polymorphisms of 12 makers used in the $F_{3:4}$ population were firstly checked between parental lines M08 and N21, and four of them (W05111, FD01071, W07092, and W07094) were effective (Fig. 2c). Then, marker FD01071 was used for linkage analysis, confirming the presence of underlying locus at chromosome 6. Subsequently, the two flanking markers W05111 and W07094 were subjected to screen a 1594 M08 x N21 F_2 segregating population (containing the 358 individuals used for inheritance analysis), obtaining a total of 44 recombinants. Basing on the high-confident SNPs identified between M08 and N21 (Wei et al. 2019), three new polymorphic markers (W05112, W11041, and W07093), as well as the four effective markers (W05111, FD01071, W07092, and W07094 from $F_{3:4}$ segregating population), were used to genotype the 44 recombinants. Finally, the *CISP* was delimited in a 611.78 Kb region between markers W11041 and W07093 (Fig. 2c), with 14 and 2 recombinants respectively, further validating the presence of recombination suppression in this locus.

Gene annotation and haplotype block analysis of *CISP* locus

According to the watermelon reference genome (97103, V2), there are 74 predicted genes in the 611.78 Kb target region. To obtain all possible annotated genes in this region, the syntenic regions of the *ClSP* locus in reference genome of 97103 (V1, 616.82 Kb) and another watermelon reference genome Charleston Gray (V1, 656.65 Kb) were comprehensively analyzed by BLAST searches, harboring 63 predicted genes for each genome (Fig. 3; Table S2). After visually removing the redundant genes, a total of 81 predicted genes were finally obtained, including three Myb transcription factors, two HD transcription factors, one WD40 transcription factor, and one NAC transcription factor, namely as candidate No1 to No7 based on their chromosome locations (Fig. 3; Table S2). Among them, the WD40 transcription factor gene *Cla97C06G126710* (candidate No2) has been recently identified as the best candidate gene for stripe pattern trait in watermelon via GWAS approach (Guo et al. 2019). Furthermore, it is worth to note that gene *Cla97C06G126560* encoding a polygalacturonase-1 noncatalytic subunit beta protein was referred as candidate No8 in the target region, because it is the ortholog of cucumber *ist* (Table S3), which is reported to regulate the irregularly striped rind pattern (Song et al. 2019).

Interestingly, according to nucleotide polymorphisms (SNPs and indels) identified in the BSR-seq data, at least four discrete haplotype blocks (haploblocks) were characterized in *ClSP* locus (Fig. S2). The first one covering about 77 Kb, exhibited consistent allelic structures among reference genome, stripe and non-stripe pools. The second haploblock spanning approximately 65 Kb and harboring candidate No8, was heterozygous in stripe pool but homozygous in reference 97103 and non-stripe pool. Similarly, the third block (~ 312 Kb) containing five candidates (No1-No5) was largest one, which was heterozygous in stripe pool and consisted of two distinct haploblocks from homozygous reference 97103 and non-stripe pool (Fig. S2). The last haploblock harboring candidates No6 and No7 covered approximately 122 Kb genomic region. Interestingly, we found this haplotype block was identical in two pools but distant from that in reference genome. In addition, an ambiguous haplotype block (~22 Kb) was also defined due to no transcriptional reads detected in this region. Nevertheless, two distinct haplotypes were distinguished between M08 and N21 based on the DNA-seq data, with stripe line M08 identical to reference 97103 (Fig. S2). Collectively, the sequence variations of haploblock among genotypes provide valuable information for candidate gene filtering.

Gene structure analysis of candidates

Intriguingly, gene structures of candidates No1-5 were completely different from their genome annotations according to our BSR-seq data (Fig. S3). For instance, candidate No1 (*Cla97C06G126680*) was predicted to contain three exons in reference genome 97103 (V2); however, the second as well as parts of the first exon was alternatively spliced in both stripe and non-stripe pools. Candidates No2 (*Cla97C06G126710*, two exons) and No3 (*Cla97C06G126720*, two exons) were adjacently distributed on chromosome 6 in the reference genome, while these two transcripts were predictably derived from an integrated amplicon (named as candidate No2_3 hereafter) according to the BSR-seq data (Fig. S3). Moreover, the original genomic sequence of candidate No3 (two exons and one intron) was transcribed as an entire exon in re-annotated No2_3. Similarly, the independent candidates No4 (*Cla97C06G126730*,

two exons) and No5 (*Cla97C06G126740*, two exons) were also presumed to transcribed from an approximately 8 Kb transcript (designated as No4_5 afterward). Meanwhile, an extra exon was detected in the re-annotated candidate No4_5, positioning in the intergenic region between No4 and No5. Regarding as candidates No6 and No8, no visible variation was observed in their gene structures (Fig. S3). It is worth to note that no transcriptional read was mapped on candidate No7 (*Cla97C06G127070*) based on the BSR-seq data of stripe and non-stripe pools.

To confirm the gene structures described above, coding sequence (CDS) for each candidate was amplified from two bulks (primer combinations as shown in Fig. S3). Sequence analysis revealed that gene structures of candidate No6 and No8 in two pools were identical to their reference genome annotations (data not shown), consistent with the BSR-seq results. Meanwhile, no PCR product was cloned for candidate No7. However, referring to candidate No1, three alternative transcripts were detected in two pools and defined as No1.1, No1.2, and No1.3 respectively (Fig. 4a). Compared to the original structure of candidate No1, only one intron was validated with 'CT..AC' as splice junction in all three transcripts; meanwhile, the first exon was in different length and the second was short for 91 bp at its 5' part. Notably, transcripts No1.1 and No1.2 were detected in both stripe and non-stripe pools, but the No1.3 was specifically transcribed in the former pool. Additionally, transcripts No1.2 and No1.3 were predicted with premature stop codons leading to truncated proteins. Similarly, three different transcripts of candidate No2_3 were also detected in both stripe and non-stripe pools, which contained two introns with 'CT..AC' as intron splice site (Fig. 4b). The second exon varied in length among these three transcripts, and the longest one was 501 bp in transcript No2_3.1. As mentioned above, the original candidate No3 (two exons and one intron) together with additional 31bp nucleotides were confirmed to be transcribed as a longer exon in re-annotated candidate No2_3. Due to candidates No2 and No3 encoding WD40 and MYB transcription factors respectively (Fig. 3; Table S2), we tried to translate all the three transcripts of candidate No2_3 into amino acids for function prediction. Unfortunately, no intact protein was obtained (Fig. S4, No2_3.2 and No2_3.3 not shown). According to the BSR-seq data, the re-annotated candidate N04_5 was hypothesized to consist of No4, No5 and an extra exon. As expected, only one transcript was detected in both stripe and non-stripe pools (Fig. 4c), with the third extra exon harboring 115 bp nucleotides. Furthermore, the candidate No4_5 corresponding to a 326 amino acid protein is a member of class I of Knotted related homeobox transcription family (Fig. S5), exhibiting high sequence identity with gene *AtKAT6* in Arabidopsis (Table S3).

Sequence polymorphic variations of candidate genes

According to our re-sequencing data (DNA-seq of M08 and N21 lines, BSR-seq of stripe and non-stripe bulks), sequence variations of the candidate genes were analyzed among stripe and non-stripe pattern genotypes. As described above (Fig. S2), candidates No1, No2_3, and No4_5 were located in a diverse haplotype block which was homozygous in reference genome 97013, M08, N21, and non-stripe pool, but heterozygous in stripe pool. Consistently, the polymorphic sites in the CDS of these three candidates were identical in non-stripe pool and N21 contrast to the stripe lines 97103 and M08, which were heterozygous

in stripe pool (Fig. 5). For both candidate No1 and No4_5, two out of four SNP sites led to amino acid changes. Notably, the third mutant at No4_5 from base C to T resulted in the 247th amino acid glutamine (Q) instead by a premature stop codon (TGA) in non-stripe plants. The subsequent amino acid substitution analysis of corresponding nucleotide mutants in candidate No2_3 was ignored, because of no intact protein predicted with its CDS (Fig. S4). In terms of candidate No6, polymorphic sites in CDS displayed an identical haplotype among two pools and the non-stripe N21, compared to that in stripe lines 97103 and M08 (Fig. 5). Similarly, sequence variations of candidate No7 displayed a consistent haplotype in stripe lines 97103 and M08, while no transcript was detected in two pools (stripe and non-stripe pools). As the ortholog of gene *ist* responsible for the irregularly stripped rind phenotype in cucumber (Song et al. 2019), candidate No8 contained three mutant sites in N21 (non-stripe phenotype) contrast to other four genotypes. Moreover, another three heterozygous sites were also confirmed in stripe pool by BSR-seq data and gene cloning assay, with the third variation causing amino acid substitution. In contrast to sequence variations in coding sequence, much more polymorphisms were detected in the introns of all the six candidates, especially in re-annotated No4_5 (Fig. 5).

Expression analysis of candidate genes

To examine the transcriptional accumulation of each candidate during stripe formation, we randomly chose a 23-DAP fruit with typical stripes, to precisely sample peel from stripe and non-stripe rind, respectively (Fig. 6a). Consistent with the BSR-seq data and gene cloning assay, no transcriptional product of candidate No7 was detected in both stripe and non-stripe peels. Additionally, only the transcription accumulation of No4_5 was significantly reduced in non-stripe peel, compared to the fluctuant expression levels of No1, No2_3, No6, and No8 (Fig. 6b), inferring the potential function of candidate No4_5 involved in stripe formation.

Putative mechanisms underlying stripe development

To explore the possible molecular mechanism responsible for stripe formation, transcriptome profiles were performed using peels precisely sampled from stripe and non-stripe rind of two striped fruits (Fig. 6a). After filtering the raw data, approximately 4.41, 4.59, 4.41, and 4.19 Gb clean data were obtained respectively, with Q30 values higher than 92.00% and more than 96.00% reads successfully mapped on reference genome 97103 (Table 2). Following p-adjust value < 0.05 and fold change >= 2 as thresholds, 213 up-regulated and 143 down-regulated differentially expressed genes (DEGs) were identified in non-striped rind (Table S4). Notably, only three DEGs were located in the *ClSP* locus, including the HD transcription factor *Cla97C06G126740* (original candidate No4), a part of re-annotated candidate No4_5.

The GO categorization analysis revealed that 141 DEGs (86 up-regulated and 55 down-regulated) were significantly enriched in 38 GO terms (Fig. 7a; Table S5), belonging to three major categories biological process (BP, 7), cellular component (CC, 25), and molecular function (MF, 6). As the most prevalently enriched category, the majority of GO terms (21 out of 25) in cellular component (CC) finally pointed to

three GO terms, ‘photosystem I’ (GO:0009522), ‘photosystem II’ (GO:0009523), and ‘chloroplast thylakoid membrane’ (GO:0009535) (Figs. 7b and S6). To better understand the metabolic pathways involved in stripe development, KEGG enrichment analysis was performed using p-value < 0.05 as the significance cut-off, obtaining six mainly concentrated metabolic pathways: photosynthesis (ko00195) and photosynthesis-antenna proteins (ko00196), metabolic pathways (ko01100), plant-pathogen interaction (ko04626), phenylalanine metabolism (ko00360), and porphyrin and chlorophyll metabolism (ko00860) (Table S6). Interestingly, we noticed that all the DEGs enriched in energy metabolism photosynthesis (ko00195) and photosynthesis-antenna proteins (ko00196) were down-regulated in non-striped rind, including 14 genes encoding Chlorophyll a_b-binding protein and 15 genes related to photosynthesis (Table S7). Moreover, twelve genes mediating calcium-signal pathway were significantly up-regulated in non-stripe peel relative to stripes, inferring the possibly uneven distribution of calcium during stripe development.

Transcriptional factors involved in stripe formation

Based on the plant transcription factor database PlantTFDB (http://planttfdb.gao-lab.org/index.php), 38 DEGs belonging to 13 families were annotated as transcription factors, including ERF, WRKY, C2H2, and NAC (Fig. 7c). Among them, the majority (31 out of 38 DEGs) were up-regulated in non-striped peels (Table S8). As the largest detected TF family, all the 9 ERF genes were significantly increased at transcriptional level, presumably suggested the vital function of plant hormone ethylene in stripe development. Additionally, six WRKY genes, four C2H2 and NAC transcription genes, as well as other 15 differently expressed TFs demonstrated the complex regulatory network of stripe formation in watermelon.

In addition, to verify the RNA-seq results, 20 DEGs were selected to quantify their relative expression profiles using qRT-PCR (Table S9). The results confirmed that genes involved in chlorophyll a_b-binding and photosynthesis processes were down-regulated in non-stripes relative to stripes (Fig. 8). However, the five genes related to calcium ion binding and ERF TFs were up-regulated in non-stripe peel. The highly consistent expression trends of 20 DEGs between RNA-seq and qRT-PCR validated the reliability of the RNA-seq data.

Discussion

As an important commercial trait for watermelon, rind appearance is the most intuitive evaluation standard for consumption, which preferentially affects the choice of customers (Kim et al. 2015; Park et al. 2016). Rind types as the major objectives in watermelon breeding, have been further divided as the foreground stripe patterns and background rind colors. The former is diverse and characterized by the margin of stripes (well-defined, medium, and diffuse), the width of stripes (very narrow to very broad), the intensity of stripe coloration (unicolored, bicolored, marbled, or only vein), the conspicuousness of stripes (inconspicuous to very strong), while the latter also presents in various phenotypes (dark green, medium

green, light green, gray, white and yellow) (Kim et al. 2015; Lou and Wehner 2016; Park et al. 2016). To date, several complex genetic models have been raised to explain the inheritance of rind pattern and fruit color in watermelon, e.g., the three alleles at a single locus *G* (dark green rind color) > g^s (stripes) > *g* (light green rind) (Poole 1944; Weetman 1937), the three independent loci *S*, *D*, and *Dgo* on chromosomes 6, 8, and 4 for foreground stripe pattern, depth of rind color, and background rind color respectively (Park et al. 2016), the five alleles at *g* locus *G* (dark green rind color) > g^W (wide stripe) > g^M(medium stripe) > g^N (narrow stripe) > *g* (light green rind color) (Lou and Wehner 2016). Using Jubille-type line 'TS34' (JT type, dark green stripes with medium width and sharp margin) from South Korea and Crimson-type line 'Arka Manik' (CT type, light green stripes with broad width and diffused margin) from India as materials, nine different stripe patterns were observed among the progenies of F_2 population, inferring the rind stripe pattern inherited by quantitative trait loci (QTLs) (Kim et al. 2015). Subsequently, the genetic locus responsible for JT stripe pattern was finally anchored on chromosome 6 from 24.03 Mb (24,030 Kb) to 26.32 Mb (26,317 Kb) according to the reference genome 97103 (V1). Consistently, using another two inbred lines '01' (JT type, standard green with stripes) and '09' (standard green without stripes), an overlapping region harboring *S* locus which is also assumed to control JT stripe pattern, was delimited from 21.78 Mb (21,778 kb) to 25.77 Mb (25,767 kb) on chromosome 6 (HeeBum et al. 2015; Park et al. 2016). In this current study, the objective stripe pattern in the two segregating populations showed highly similar to JT phenotype (HeeBum et al. 2015; Kim et al. 2015), which was also controlled by a single dominant gene (Figs. 1, S1; Table 1). Based on reference genome 97103 (V1), the causal gene was finally narrowed down in a 616.82 Kb region between markers W11041 and W07093 from 25.67 Mb (25,666 Kb) to 26.28 Mb (26,283 Kb) (Fig. 3), exhibiting an obvious recombination suppression (approximately 2880 individuals used in this mapping strategy). Compared to the *S* locus, approximately 100 kb overlapping region was identified, which was mainly included in the first haplotype block exhibiting no sequence polymorphisms between stripe and non-stripe pools (Fig. S2). Therefore, we presumably speculated that the target gene in this study was differed from the *S* locus, and subsequently designated as *ClSP*. Apparently, different materials considerably lead to the ambiguous conclusions during map-based cloning strategy, e.g., the stripe epicarp (*st*) in melon line 'Dulce' cv (*Cucumis melo var. reticulatus*) is recessive to the non-striped phenotype, and is located to the linkage group III(6) (Danin-Poleg et al. 2002); however, the stripes on green rind in line 'X010' is dominant over the white rind without stripes of 'M1-113', and the responsible gene *st3* is anchored on chromosome 4 (Liu et al. 2019). Given the complex mechanism of stripe formation in watermelon, further genetic analyses are necessary in future, such as the allelism examination between stripe pattern related loci *ClSP* and *S*.

For horticultural crops, fruit morphological features determine their market acceptance, which are important to both breeders and consumers. To date, several transcription factors have been reported to be responsible for stripe pattern in plants. For example, in apple and pear, the methylation levels of *MYB10* promoter region are associated with the red and green stripes (Qian et al. 2014; Telias et al. 2011). Similarly, the methylation of the *TAGL1* promoter, a MADS-box transcription factor, reduces its expression and leads to green stripes of fruit in tomato (Liu et al. 2020). In cucurbit crops, a MYC2-like transcription factor *MELO3C003412* has been listed as one of the most likely candidates responsible for striped rind in

melon (Liu et al. 2019). The irregularly stripped rind underlying gene *ist* in cucumber is predicted to be *Csa1G005490*, encoding a polygalacturonase-1 noncatalytic subunit beta protein (PG1β) known to be involved in fruit softening (Song et al. 2019). Considering the distribution characteristics of stripes on watermelon fruit, we focused on analyzing the transcription factors in the *ClSP* mapping interval, and obtained seven candidates namely No1 to No7 (Fig. 3). Additionally, gene *Cla97C06G126560*, the ortholog of gene *ist*, was considered as candidate No8. Given that there were three striped genotypes (reference 97103, M08, and stripe pool) and two non-striped genotypes (N21 and non-stripe pool), sequence polymorphisms in *ClSP* mapping region were comprehensively compared among these five genotypes based on RNA-seq and DNA-seq data. Interestingly, at least four discrete haplotype blocks were characterized in *ClSP* locus (Fig. S2), with candidates No8, No1-5, and No6-7 located in three independent haploblocks. According to RNA-seq data, gene structures of candidates No1-5 predictably varied from their reference annotations, which were subsequently verified by TA cloning strategy (Figs. 4, S3). Sequence analyses revealed that at least four SNP mutants were detected in CDS of each candidate (Fig. 5). However, only the candidate No4_5 exhibited different transcriptional accumulations between striped and non-striped peels (Fig. 6b). Using GWAS approach, the WD40-repeat gene *Cla97C06G126710* showed the strongest signal associated with rind stripe in watermelon (Guo et al. 2019), which was listed as candidate No2 in this study, a part of re-annotated candidate No2_3 (Fig. 5; Table S2). Comparative sequence analyses suggested that the candidate No2_3 was a probable pseudogene with no complete set of open reading frames (Fig. S4), which exhibited similar expression accumulations between stripe and non-stripe peels (Fig. 6b). Collectively, the re-annotated candidate No4_5 was considered as the greatest potential gene for *ClSP*, presumably owing to its significantly reduced expression level in non-stripes, together with the premature stop codon (TGA) leading to truncated proteins in non-stripe plants. On the other hand, except for these eight candidates, there were too many genes in the *ClSP* locus (Table S2). We could not accurately assign the causal gene responsible for stripe formation in watermelon. Therefore, transformation experiments are needed for functional verification of candidate No4_5 in future research.

In this study, 356 DEGs were detected between stripe and non-stripe peels by RNA-seq, including 213 up-regulated and 143 down-regulated in the latter (Table S4). The GO and KEGG enrichment analyses revealed that a series of items related to photosynthesis and chloroplast structure were significantly enriched (Fig. 7; Tables S5, S6). Photosynthesis is reported to be associated with the chloroplast formation and pigments synthesis (Song et al. 2019). The down-regulation of photosynthesis related genes can consequently effect the chloroplast thylakoid development. Consistently, DEGs enriched in the photosynthesis and chloroplast development were all down-regulated in non-stripes relative to stripes (Tables S5, S6). Recently, calcium has been validated to spatially coordinate with red stripes on pear fruit (Zhai et al. 2019). Interestingly, 12 DEGs involved in calcium ion binding process were all significantly increased in non-stripe peel (Table S7), inferring the possible uneven distribution of calcium during stripes formation in watermelon. Transcription factors responsible for peel color and stripe pattern formation have been widely studied in crops, e.g., the transcription complex *MYB-bHLH-WD40* in flavonoid biosynthetic pathway (Xu et al. 2015), *MYB10* in apple and pear associated with stripes (Qian et al. 2014;

Telias et al. 2011), *MYB36* in cucumber responsible for peel color (Hao et al. 2018), a MADS-box transcription factor gene *TALG1* related to stripes in tomato (Liu et al. 2020). Here, we identified 38 differently expression transcription factors belonging to 13 gene families, such as ERF, WRKY, C2H2, and NAC, and the majority (31 out of 38) were up-regulated in non-striped peels (Fig. 7c; Table S8). Most strikingly, expression levels of nine ERFs (ethylene responsive factor) were significantly enhanced in non-stripes, suggested that the plant hormone ethylene may functions in the stripe formation in watermelon.

Declarations

Acknowledgments

This work was supported by funding from the National Natural Science Foundation of China [Grant No. 31701939] and the National Natural Science Foundation of Shaanxi Province, China [No. 2019JQ-324].

Author contribution statement

CW designed the experiment and wrote the manuscript, with help from XZ. ZY performed the major experiments and analyzed the data. RM, DC, and XY participated in DNA extraction and phenotypic record. YH and CW contributed to gene cloning and sequence analysis. XP and LY participated in RNA extraction. All authors have read and approved the final manuscript.

Conflict of interest

The authors declare that there is no confict of interest.

References

Danin-Poleg Y, Tadmor Y, Tzuri G, Reis N, Hirschberg J, Katzir N (2002) Construction of a genetic map of melon with molecular markers and horticultural traits, and localization of genes associated with ZYMV resistance. Euphytica 125:373-384

Gama R, Santos C, Dias R, Alves J, Nogueira T (2015) Microsatellite markers linked to the locus of the watermelon fruit stripe pattern. Genet Mol Res 14:269-276

Grumet R, Colle M (2016) Genomic analysis of cucurbit fruit growth. Genetics and Genomics of Cucurbitaceae. Springer, pp 321-344

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51-58

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

Gusmini G, Wehner TC (2006) Qualitative inheritance of rind pattern and flesh color in watermelon. J Heredity 97:177-185

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

HeeBum Y, SungWoo P, GungPyo L, SunCheol K, YongKwon K (2015) Linkage analysis of the three loci determining rind color and stripe pattern in watermelon. Korean J Hortic Sci 33:559-565

Kim H, Han D, Kang J, Choi Y, Levi A, Lee GP, Park Y (2015) Sequence-characterized amplified polymorphism markers for selecting rind stripe pattern in watermelon (*Citrullus lanatus* L.). Hortic Environ Biote 56:341-349

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359

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

Liu G, Li C, Yu H, Tao P, Yuan L, Ye J, et al (2020) *GREEN STRIPE,* encoding methylated *TOMATO AGAMOUS-LIKE 1*, regulates chloroplast development and Chl synthesis in fruit. New Phytol 228:302-317

Liu L, Sun T, Liu X, Guo Y, Huang X, Gao P, Wang X (2019) Genetic analysis and mapping of a striped rind gene (*st3*) in melon (*Cucumis melo* L.). Euphytica 215:20

Lou L, Wehner TC (2016) Qualitative inheritance of external fruit traits in watermelon. HortScience 51:487-496

Lv J, Fu Q, Lai Y, Zhou M, Wang H (2018) Inheritance and gene mapping of spotted to non-spotted trait gene *CmSp-1* in melon (*Cucumis melo* L. var. chinensis Pangalo). Mol Breeding 38:105

Paris HS (2003) Genetic control of irregular striping, a new phenotype in *Cucurbita pepo*. Euphytica 129:119-126

Paris HS (2009) Genes for “reverse” fruit striping in squash (*Cucurbita pepo*). J Heredity 100:371-379

Park S-w, Kim K-T, Kang S-C, Yang H-B (2016) Rapid and practical molecular marker development for rind traits in watermelon. Hortic Environ Biote 57:385-391

Park Y, Cho S (2012) Watermelon production and breeding in South Korea. Isr J Plant Sci 60:415-423

Poole C (1944) Genetics of cultivated cucurbits. J Heredity 35:122-128

Qian M, Sun Y, Allan AC, Teng Y, Zhang D (2014) The red sport of 'Zaosu' pear and its red-striped pigmentation pattern are associated with demethylation of the *PyMYB10* promoter. Phytochemistry 107:16-23

Rhodes B, Zhang X (1995) Gene list for watermelon. Cucurbit Genet Coop

Song M, Zhang M, Cheng F, Wei Q, Wang J, Davoudi M, Chen J, Lou Q (2019) An irregularly striped rind mutant reveals new insight into the function of PG1beta in cucumber (*Cucumis sativus* L.). Theor Appl Genet 133:371-382

Su W, Tao R, Liu W, Yu C, Yue Z, He S, et al (2019) Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. Plant Biotechnol J 18:479-490

Telias A, Lin-Wang K, Stevenson DE, Cooney JM, Hellens RP, Allan AC, Hoover EE, Bradeen JM (2011) Apple skin patterning is associated with differential expression of *MYB10*. BMC Plant Biol 11:93

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178-192

Weetman LM (1937) Inheritance and correlation of shape, size and color in the watermelon, Citrullus vulgaris Schrad. Iowa Agriculture and Home Economics Experiment Station Research Bulletin 20:1

Wehner T, Pitrat M (2008) Overview of the genes of watermelon p. 79–89. In: Pitrat, M. (ed.). Cucurbitaceae 2008. Proc. IXth EUCARPIA meeting on genetics and breeding of Cucurbitaceae, Il'institut National de la Recherche Agronomique, Avignon, France, 21–24 May 2008.

Wei C, Chen X, Wang Z, Liu Q, Li H, Zhang Y, et al (2017) Genetic mapping of the *LOBED LEAF 1* (*ClLL1*) gene to a 127.6-kb region in watermelon (*Citrullus lanatus* L.). PloS One 12:e0180741

Wei C, Zhu C, Yang L, Zhao W, Ma R, Li H, et al (2019) A point mutation resulting in a 13 bp deletion in the coding sequence of *Cldf* leads to a GA-deficient dwarf phenotype in watermelon. Hortic Res 6:132

Wu S, Wang X, Reddy U, Sun H, Bao K, Gao L, et al (2019) Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. National Plant Germplasm System watermelon collection. Plant Biotechnol J 17:2246-2258

Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci 20:176-185

Yan C, An G, Zhu T, Zhang W, Zhang L, Peng L, et al (2019) Independent activation of the *BoMYB2* gene leading to purple traits in Brassica oleracea. Theor Appl Genet 132:895-906

Zhai R, Wang Z, Yang C, Lin-Wang K, Espley R, Liu J, et al (2019) *PbGA2ox8* induces vascular-related anthocyanin accumulation and contributes to red stripe formation on pear fruit. Hortic Res 6:137

Zhang Z, Zhang Y, Sun L, Qiu G, Sun Y, Zhu Z, et al (2018) Construction of a genetic map for *Citrullus lanatus* based on CAPS markers and mapping of three qualitative traits. Sci Hortic 233:532-538

Tables

Table 1 Genetic inheritance analysis of stripe patterns in two segregating population

Generation	Total number	Stripes	Non-stripes	Expected ratio	χ^2-value	P-value
$F_{3:4}$ population	1286	987	299	3:1	2.01	0.15
M08	30	30	0			
N21	30	0	30			
F_1	20	20	0			
F_2 population	358	281	77	3:1	2.15	0.13

χ^2 (0.05, 1) =3.84

Table 2 Detailed information of RNA-seq data generated in this study

	BSA+RNA-seq		RNA-seq			
	Stripe pool	Non-stripe pool	Stripe_1	Stripe_2	Non-stripe_1	Non-stripe_2
Clean Reads Num	43930646	51123222	29379434	29383944	30578246	27928474
Clean Data(Gb)	6.59	7.67	4.41	4.41	4.59	4.19
Q30 Percentage	91.97%	92.69%	92.88%	93.03%	93.55%	93.53%
GC%	45.32%	44.91%	43.29%	43.83%	43.53%	44.22%
Mapped Reads	39799576 (90.6%)	48001240 (93.89%)	28269491 (96.22%)	28289745 (96.28%)	29542475 (96.61%)	27091935 (97.0%)
Uniquely Mapped Reads	39130276 (89.07%)	47169134 (92.27%)	27716863 (94.34%)	27657251 (94.12%)	28964970 (94.72%)	26578646 (95.17%)

Figures

Figure 1

Two distinct stripe patterns of watermelon. a The dark green stripes with clear margins on a standard green background, designated as stripe pattern. b The netted reticulations on the standard green rind, designated as non-stripe pattern. The two different stripe patterns displayed high similar to phenotypes described in recent publishes (HeeBum et al. 2015; Park et al. 2016). The dynamic morphological features of stripe pattern were recorded at different fruit developmental stages. The magnified rinds represented tissues for BSR-seq. Bold letters ‘B’ and ‘D’ referred as ‘Day before anthesis’ and ‘Days after pollination’, respectively.

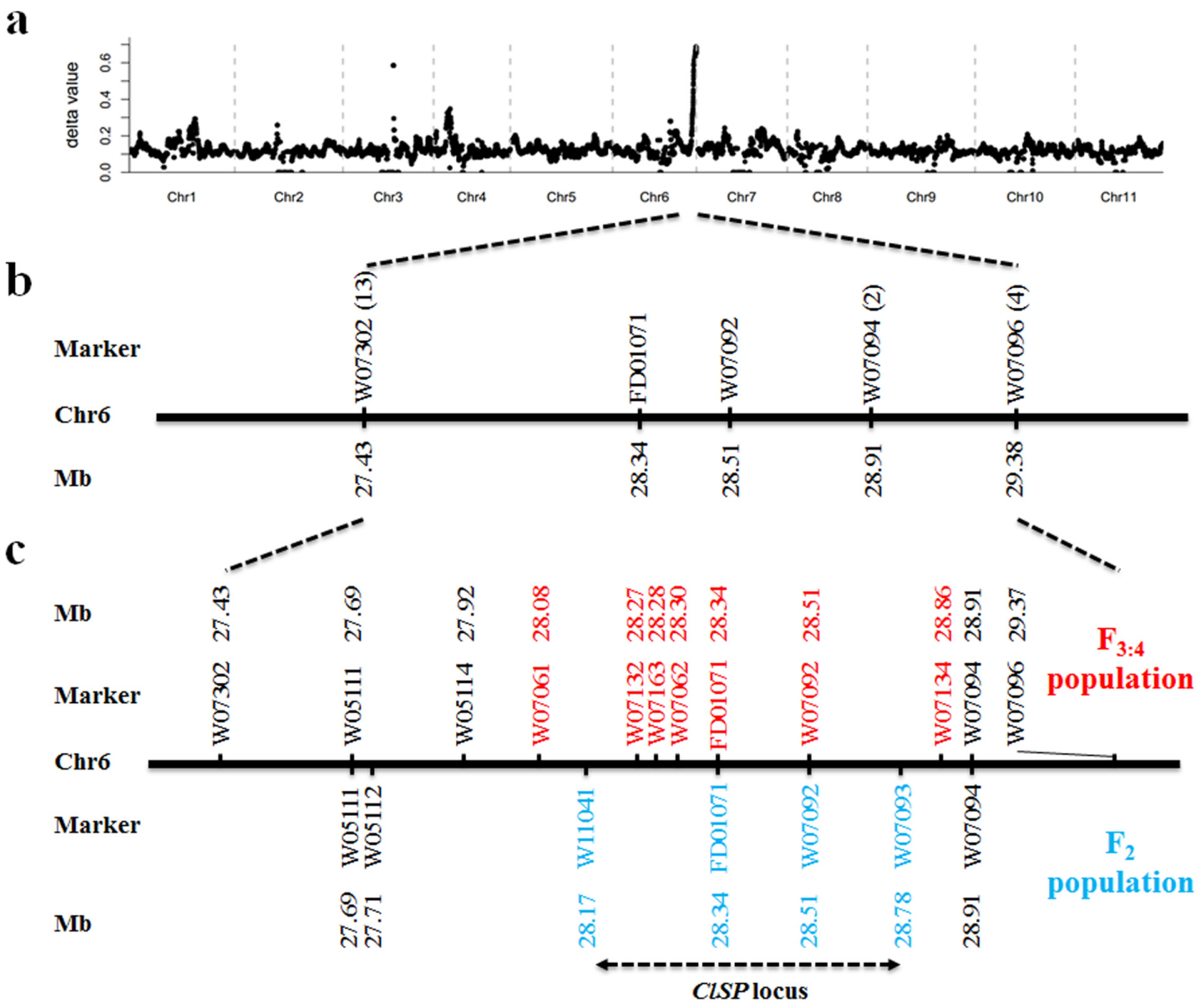

Figure 2

Genetic mapping of ClSP locus. a The average values of delta SNP index (y-axis) plotted along the eleven chromosomes (x-axis) of watermelon. A predominant peak located on chromosome 6. b The ClSP locus was primarily delimited between markers W07302 and W07096 on chromosome 6, using the F3:4 segregating population (n = 119). c The ClSP locus was finally narrowed down in a 611.78 Kb region using two different segregation populations, the F3:4 population (n = 1286) and M08 x N21 F2 population (n= 1594).

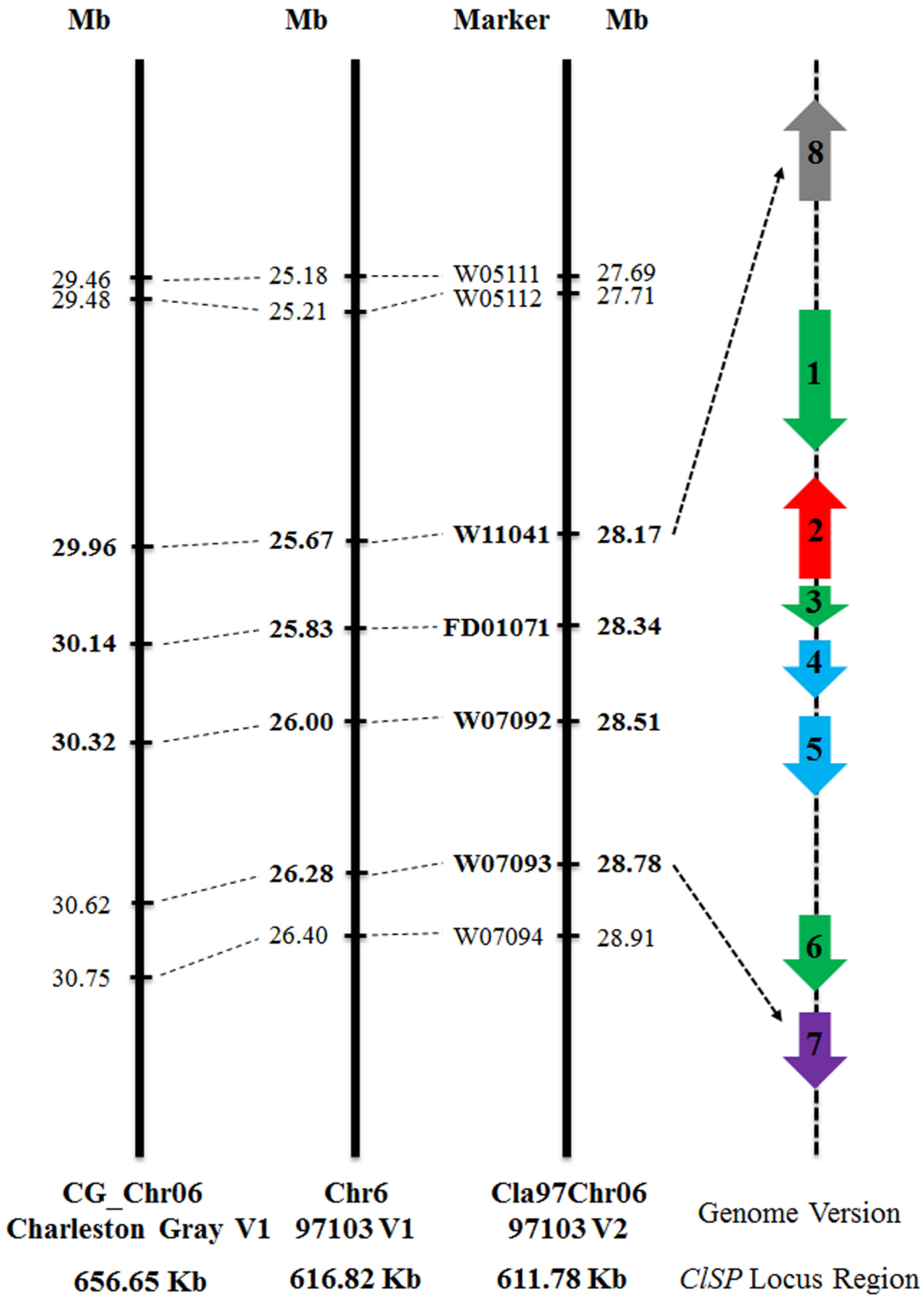

Figure 3

Candidate gene analysis in the ClSP locus. Using BLAST program, the polymorphic markers were comparatively mapped on reference genomes 97103 (version 1) and Charleston Gray (version 1). Taken together, seven transcription factors (TFs) were detected in this region, i.e., two MYB TFs (No1 and No3 marked in green), one WD40 TF (No2 in red), two HD TFs (No4 and No5 in blue), and one NAC TF (No7 in

purple). Additionally, the ortholog gene of cucumber ist responsible for irregularly striped rind phenotype was considered as candidate No8 and marked in gray. The arrows represent the transcription directions.

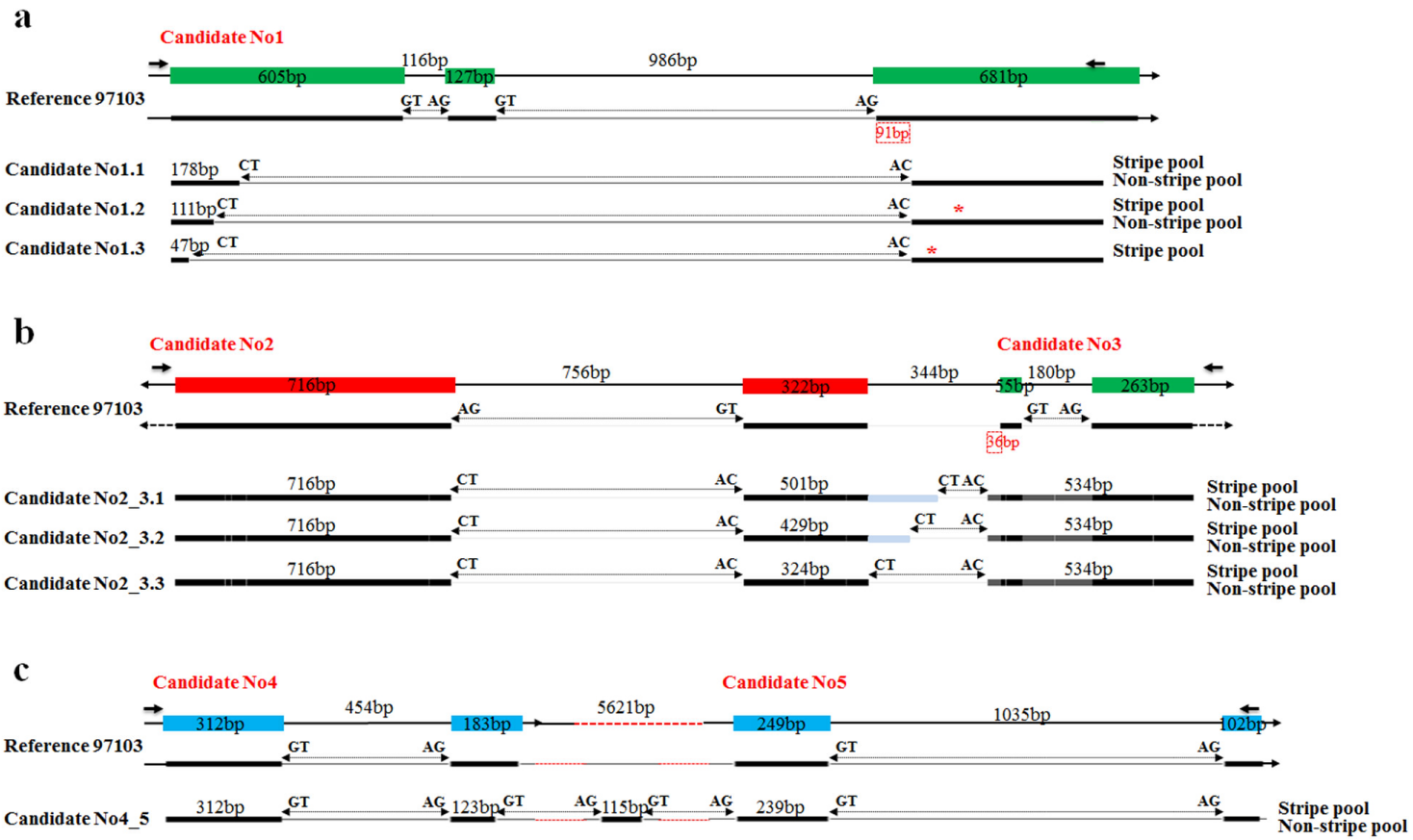

Figure 4

Re-annotated gene structures of candidates No1-No5. Their original structures in reference genome 97103 (V2) were marked in colors, with arrows representing primer positions for gene cloning. a Schematic representation of candidate No1. Three alternative transcripts were detected in stripe and non-stripe pools, named as No1.1, No1.2 and No1.3 consisting of two partial exons. Among them, transcript No1.3 was only detected in stripe pool. b Gene structure of re-annotated candidate No2_3. The two adjacent distribution gene No2 and No3 were confirmed to transcribe from the re-annotated No2_3, with three different transcripts (namely No2_3.1, No2_3.2, and No2_3.3) in both stripe and non-stripe pools. The second exon was in different length, and the last harbored genomic sequence of No3 and additional 36 bp nucleotides. c Re-annotated gene structure of candidate No4_5 consisting of No4, No5, and the third extra exon.

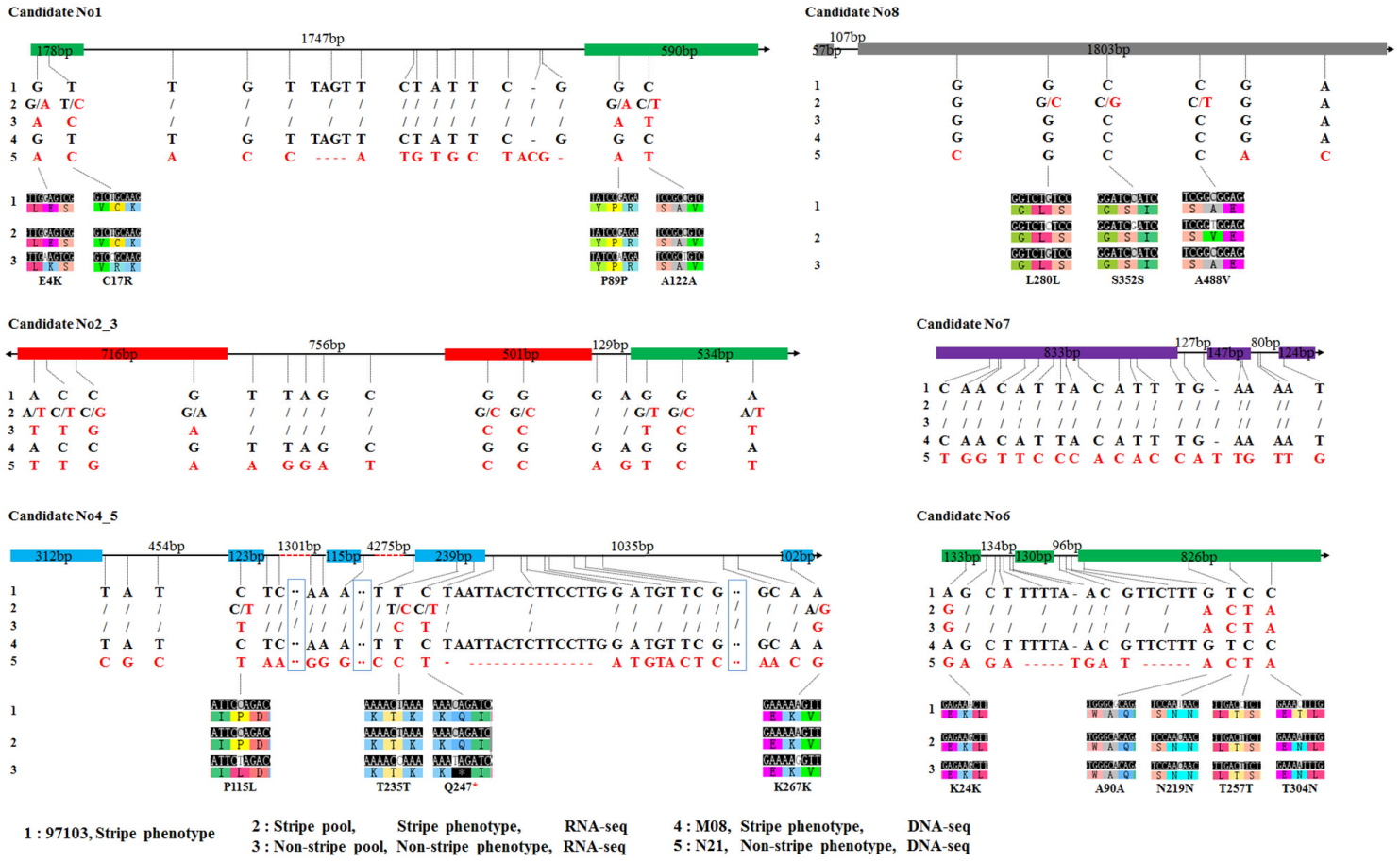

Figure 5

Sequence analyses of candidates among reference 97103, M08, N21, stripe and non-stripe pools. Schematic diagrams for each candidate were presented. Polymorphic variations were marked in red. Amino acid substitutions corresponding to SNPs identified among reference 97013, stripe and non-stripe pools were also displayed in candidates No1, No4_5, No6, and No8. Notably, a premature stop codon in re-annotated No4_5 led to truncated proteins in non-stripe pool.

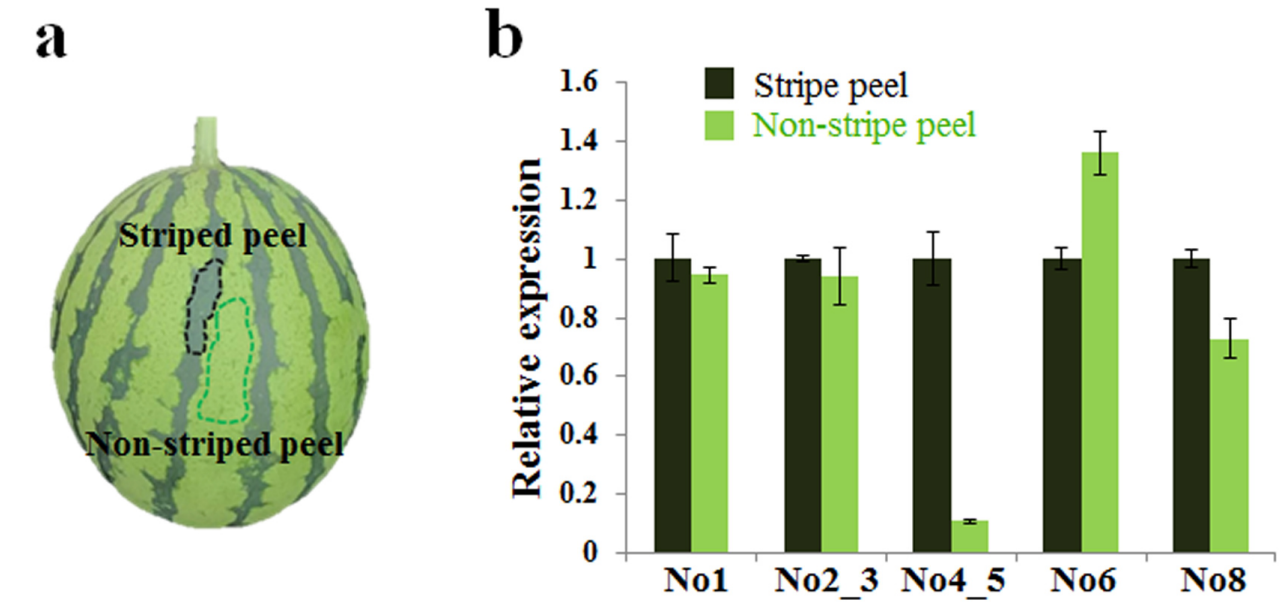

Figure 6

Expression profiles of candidates between striped and non-striped peel. a Regular distribution of stripes over watermelon fruits. Striped and non-striped peels were precisely sampled to extract total RNA for expression profiles. b Transcription levels of five candidates between stripes and non-stripes. Due to no transcript detected in both two samples, candidate No7 was not displayed here. Data are represented as means ± SD.

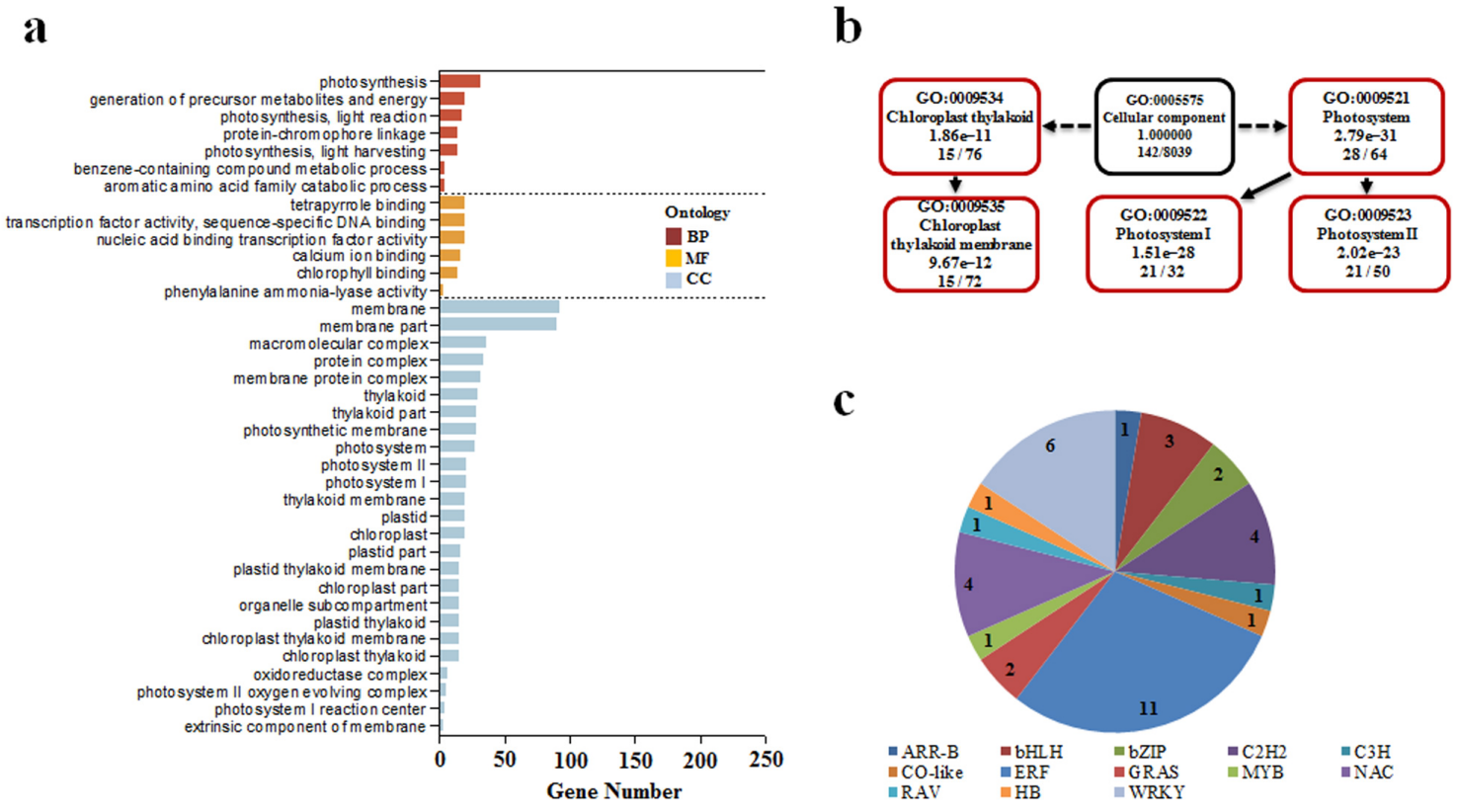

Figure 7

Enrichment analysis of DEGs. a Significant GO terms in three categories. BP: biological process, MF: molecular function, CC: cellular component. b Brief schematic diagram of GO terms in cellular component pointing to three GO terms, ‘photosystem I’ (GO:0009522), ‘photosystem II’ (GO:0009523), and ‘chloroplast thylakoid membrane’ (GO:0009535). c Transcription factors annotation of DEGs.

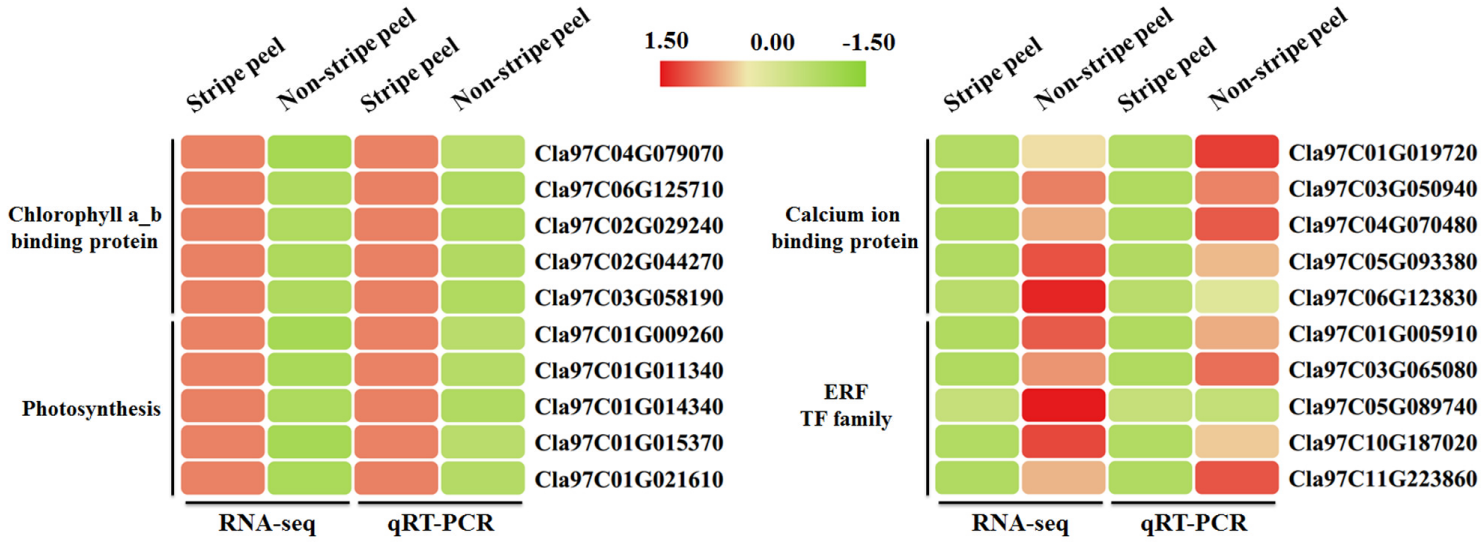

Figure 8

Validation of RNA-seq results via qRT-PCR. The relative expressions of 20 DEGs were represented in Heatmap, including five genes encoding chlorophyll a_b-binding protein, five involved in photosynthesis, five genes related to calcium ion binding process, and five ERF TFs. Data are presented as means ± SD.

Supplementary Files

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

- FigS1.tif
- FigS2.tif
- FigS3.tif
- FigS4.tif
- FigS5.tif
- FigS6.tif
- SupplementaryTables.xlsx