Identification of CaAN3 as a fruit-specific regulator of anthocyanin biosynthesis in pepper (Capsicum annuum)

Jinyoung Byun
  Seoul National University

Tae-Gun Kim
  Seoul National University

Joung-Ho Lee
  Seoul National University

Ning Li
  Hubei Academy of Agricultural Sciences

Soyoun Jung
  Seoul National University

Byoung-Cheorl Kang (✉ bk54@snu.ac.kr)
  Seoul National University  https://orcid.org/0000-0002-7606-7258

Research Article

Keywords: Anthocyanin, Fine-mapping, Capsicum, R2R3 MYB transcription factor, Transient overexpression, Virus-induced gene silencing.

Posted Date: February 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1283047/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Identification of CaAN3 as a fruit-specific regulator of anthocyanin biosynthesis in pepper (Capsicum annuum)

Jinyoung Byun1, Tae-Gun Kim2, Joung-Ho Lee1, Ning Li3, Soyoung Jung1, Byoung-Cheol Kang1*

1 Department of Agriculture, Forestry, and Bioresources, Research Institute of Agriculture and Life Sciences, Plant Genomics Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea

2 Interdisciplinary Program in Agricultural Genomics, Department of Agriculture, Forestry, and Bioresources, Research Institute of Agriculture and Life Sciences, Plant Genomics Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea

3 Cash Crops Research Institute, Hubei Academy of Agricultural Sciences, Wuhan City, Hubei Province, 430064, China

* Correspondence:
Byoung-Cheol Kang (bk54@snu.ac.kr)
Abstract

The key regulatory gene CaAN2 encodes an R2R3 MYB transcription factor that regulates anthocyanin biosynthesis in various tissues in pepper (*Capsicum annuum*). However, CaAN2 is not expressed in certain pepper accessions showing fruit-specific anthocyanin accumulation. In this study, we identified the novel locus CaAN3 as a regulator of fruit-specific anthocyanin biosynthesis, using an F$_2$ population derived from a hybrid cultivar with purple immature fruits and segregating for CaAN3. We extracted total RNA, assembled two RNA pools according to fruit color, and carried out bulked segregant RNA sequencing. We aligned the raw reads to the pepper reference genome Dempsey and identified 6,672 significant single nucleotide polymorphisms (SNPs) by calculating the Δ(SNP-index) between the two pools. We then conducted molecular mapping to delimit the target region of CaAN3 to the interval 184.6–186.4 Mbp on chromosome 10. We focused on Dem.v1.00043895, encoding an R2R3 MYB transcription factor, as the strongest candidate gene. Sequence analysis revealed four insertion/deletion polymorphisms in the promoter region of the green CaAN3 allele. We employed virus-induced gene silencing and transient overexpression assays to characterize the function of the candidate gene. When Dem.v1.00043895 was silenced in pepper, anthocyanin accumulation decreased in the pericarp, while the transient overexpression of *Dem.v1.00043895* in *Nicotiana benthamiana* leaves resulted in the accumulation of anthocyanins around the infiltration sites. These results showed that Dem.v1.00043895 is CaAN3, an activator of anthocyanin biosynthesis in pepper fruits.
Keywords
Anthocyanin, Fine-mapping, Capsicum, R2R3 MYB transcription factor, Transient overexpression, Virus-induced gene silencing.

Key Message
The novel gene CaAN3 encodes an R2R3 MYB transcription factor that regulates fruit-specific anthocyanin accumulation.
Introduction

Anthocyanins are plant secondary metabolites derived from the flavonoid biosynthesis pathway (Koes et al. 2005; Tanaka et al. 2008). They are among the most abundant pigments found in flowers and fruits, imparting red, blue, and purple coloration. Anthocyanins exert core functions in plants such as attracting pollinators and protecting against biotic and abiotic stresses (de Pascual-Teresa et al. 2010; Jimenez-Garcia et al. 2013). Anthocyanins offer benefits for human health, including anti-inflammatory and anti-carcinogenic properties, as well as the potential to prevent cardiovascular disease and to control obesity or diabetes (Cassidy et al. 2011; He & Monica Giusti 2010; Khoo et al. 2017; Lin et al. 2017). Since anthocyanins have both aesthetic value and health benefits, breeding cultivars of various crops with increased anthocyanin contents has become a priority (Allan & Espley 2018).

The anthocyanin biosynthesis pathway has been described in multiple horticultural crops, as well as in the model plants Arabidopsis and petunia (Jaakola 2013; Liu et al. 2018; Pelletier et al. 1997; Tsukaya et al. 1991). It is a branch of the flavonoid biosynthesis pathway and consists of two types of gene: structural genes that encode the enzymes directly participating in anthocyanin biosynthesis, and regulatory genes that regulate the expression of the structural genes (Gonzali et al. 2009). Structural genes are further classified into early biosynthetic genes (EBGs) that synthesize dihydroflavonols, and late biosynthetic genes (LBGs) that catalyze the conversion of leucoanthocyanidins to anthocyanidins (Dubos et al. 2010; Petroni & Tonelli 2011).

The activation of anthocyanin structural genes is coordinated, and transcription factors (TFs) that directly regulate their expression have been identified in several species (Jaakola 2013). The core of the transcriptional regulation of anthocyanin structural genes relies on the interaction of a DNA-binding R2R3 MYB TF, with a MYC-like basic helix-loop-helix (bHLH) and WD40-repeat proteins (Koes et al. 2005), forming a MYB-bHLH-WD40 (MBW) complex.
The MYB TF is the most crucial component of this complex and can induce anthocyanin accumulation by itself (Hichri et al. 2010; Kiferle et al. 2015; Spelt et al. 2000; Stracke et al. 2007).

The Solanaceae family includes widely cultivated horticultural crops such as tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), potato (*Solanum tuberosum*), and pepper (*Capsicum annuum*). Several crop species of the Solanaceae accumulate anthocyanins (Dhar et al. 2015). Due to the commercial importance of Solanaceous crops, substantial work has explored the regulatory mechanisms behind anthocyanin accumulation (Liu et al. 2018), mainly in edible parts such as fruits or tubers. In tomato, *SlANT1* and *SlAN2* encode R2R3 MYB activators, causing anthocyanin accumulation in vegetative tissues and fruits (Povero et al. 2011; Schreiber et al. 2012). *SlAN2-like*, encoding another R2R3 MYB TF whose expression is driven by a fruit-specific promoter, acts as a master regulator in anthocyanin biosynthesis, activating the transcription of both structural and regulatory genes (Sun et al. 2019). In eggplant, the functional R2R3 MYB genes *SmMYB1* and *SmMYBc* induce the expression of structural genes and lead to the accumulation of anthocyanins in fruits (Gisbert et al. 2016; Stommel & Dumm 2015; Zhang et al. 2014). In potato, *StAN1*, *StMYBA1*, and *StMYB113* encode R2R3 MYB TFs that activate anthocyanin biosynthesis and are highly expressed in purple tubers; their expression levels are positively correlated with the transcription levels of structural genes and with anthocyanin concentration (Liu et al. 2016; Strygina et al. 2019).

In pepper, relatively few studies have addressed regulatory genes. Of the few known examples, CaAN2 is a R2R3 MYB TF that regulates anthocyanin biosynthesis (Borovsky et al. 2004). The accumulation of anthocyanins is associated with the insertion of a non-long terminal repeat (LTR) retrotransposon into the CaAN2 promoter region. Various tissues including fruits (only at the immature stage) and flowers show purple pigmentation when plants carry a functional CaAN2 allele. In the absence of transcription from the structural genes, immature
fruits remain green (Jung et al. 2019; Ohno et al. 2020). However, the alleles at CaAN2 fail to
fully explain all instances of purple pigmentation in pepper fruits, as certain pepper accessions
carrying a non-functional CaAN2 allele still show full purple pigmentation specifically in
immature fruits (Jung 2019). By contrast, other accessions with a functional CaAN2 allele
exhibit purple pigmentation in flowers, leaves, and fruits.

Bulked segregant analysis (BSA) can be used to identify markers linked to any specific
gene or genomic region using two pools of DNA samples. Each pool contains individuals that
share an identical target trait or genomic region but segregate independently at all other genomic
regions (Michelmore et al. 1991). Bulked segregant RNA sequencing (BSR-seq) is a modified
BSA technique that employs RNA instead of DNA, making it possible to efficiently isolate
genes even in populations for which no polymorphic markers have been previously identified
(Liu et al. 2012). Transcriptome deep sequencing (RNA-seq) is a widely adopted application
of next-generation sequencing technologies and allows the comparative quantification of gene
expression based on a phenotype of interest (Marioni et al. 2008). RNA-seq can inform on
variation in coding regions such as single nucleotide polymorphisms (SNPs), which can also
be used as genetic markers (Chepelev et al. 2009).

In this study, we identified the novel locus CaAN3 regulating fruit-specific anthocyanin
accumulation using BSR-seq. CaAN3 was previously proposed to regulate fruit-specific
anthocyanin biosynthesis (Jung 2019). We previously developed a segregating F2 population
by crossing two pepper accessions with the non-functional CaAN2 allele, Capsicum annuum
‘MAB1’ with green immature fruits and C. annuum ‘MAB2’ with fruit-specific purple
pigmentation. We observed a 3:1 segregation ratio for purple fruits, indicating that CaAN3 is a
single dominant locus. Genetic mapping suggested that CaAN3 is located on chromosome 10,
but fine-mapping was not possible due to the large size of a region with no recombination in
the mapping interval (Jung 2019). To fine-map CaAN3, we developed an F2 population from a
hybrid bell pepper cultivar with a non-functional CaAN2 allele and with purple pigmentation in fruits. The F$_2$ population segregated for purple and green immature fruits, allowing mapping of the responsible gene. Here, we report on the fine-mapping and identification of the CaAN3 locus. We validated the identity of the candidate gene by conducting virus-induced gene silencing (VIGS) and transient overexpression. Finally, we explored the expression mechanism of CaAN3 using diverse pepper accessions with purple pigmentation in fruits and/or flowers.
Materials and methods

Plant materials

The two *C. annuum* accessions MAB2 and MAB1 were obtained from Asia Seed Co., Ltd. (Incheon, Korea) and used as controls for phenotypic and genotypic analyses. MAB2 has purple immature fruits, but green leaves and stems, while MAB1 has green immature fruits, leaves, and stems (Fig. 1a). The fruit-specific purple pigmentation line MAB2 was used in subsequent experiments after the candidate gene was identified. To map the *CaAN3* locus, the purple hybrid bell pepper cultivar ‘Salad Piment Purple’ from Takii Seed Co., Ltd. (Kyoto, Japan) was purchased from the seed market to generate a segregating F2 population. An F2 population of 243 individuals was used for BSR-seq analysis. In addition, 13 pepper accessions obtained from the National Institute of Agricultural Science (Wanju, Korea) were used to validate the identity of *CaAN3*.

Nucleic acid extraction

Genomic DNA was extracted from fresh young leaves by the cetyltrimethyl-ammonium bromide method (Lee et al. 2017). The extracted genomic DNA was dissolved in 1× Tris-HCl EDTA buffer and then diluted to a concentration of 50 ng/µL with triple-distilled water. Pericarp tissues of immature fruits were used for BSR-seq and expression analyses. Total RNA was extracted from pericarp from immature fruits using the MG RNAzol Kit (MGmed, Seoul, Korea) according to the manufacturer’s instructions.

BSR-seq

An equal amount of RNA was sampled from each of 18 purple and 18 green fruits derived from the F2 population and pooled to make three pools of samples from six individuals per fruit color. The TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) was
used to construct RNA-seq libraries; RNA-seq was performed at Macrogen (Seoul, Korea).

After removing adaptors and low-quality reads, raw reads were aligned to the pepper Dempsey v1.0 reference genome (unpublished), utilizing STAR version 2.7.5a (Dobin et al. 2013). Then, the alignments from purple and green RNA pools were analyzed using a quantitative trait locus (QTL) sequencing analysis pipeline (Takagi et al. 2013). SNPs were extracted from the alignment files of the two sets of pools using Samtools. SNPs were filtered and plotted using internal Perl and R scripts of the QTL-seq pipeline. The SNP-index value was computed for each pool as the number of aligned reads to the reference genome. The Δ(SNP-index) value was defined as the difference between the SNP-indices of the purple and green RNA pools.

**Development of a molecular marker for CaAN3**

The previously delimited CaAN3 target region in CM334 reference genome (Jung 2019) was used for further marker development. Based on the SNPs identified by BSR-seq, markers for high-resolution melting (HRM) were developed with an amplicon size of 100–300 bp, to detect polymorphisms in PCR amplicons. Three primers for the sequence-characterized amplified region (SCAR) marker were developed to amplify the CaAN3 promoter region spanning the insertion/deletion (InDel) polymorphism between purple and green alleles. PCR primers were designed with Primer3 (http://web.bioneer.co.kr/cgibin/primer/primer3.cgi).

HRM markers were mainly used for fine-mapping of CaAN3, and the SCAR marker was used to genotype various pepper accessions. Primers used as molecular markers are listed in Table S1.

**HRM analysis**

HRM markers were developed based on the BSR-seq results to test for polymorphisms in the developed markers across the F₂ population. HRM markers were used to genotype F₂
individuals on a Rotor-Gene 6000 real-time PCR thermocycler (Corbett Research, Sydney, Australia). Quantitative PCR (qPCR) was performed in reactions containing 2.5 μL 10× HiPi reaction buffer (Elpis Biotech, Daejeon, Korea), 2 μL 10 mM dNTPs, 0.5 μL each 10-pmol primers, 2 μL 50 ng/μL genomic DNA, 0.3 μL Taq polymerase and sterile distilled H₂O up to 20 μL. qPCR reactions consisted of 55 cycles of denaturation at 95°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 30 s. After the PCR, HRM analysis was carried out with increasing temperature of 0.1°C every min from 65°C to 95°C.

Construction of VIGS vectors

The pTRV2-LIC vectors were constructed using the ligation independent cloning (LIC) method as previously described (Kim et al. 2017). Partial coding sequences (200–400 bp) of the CaAN3 candidate gene were amplified with LIC adapter primers. The resulting purified PCR amplicon was treated with T4 DNA polymerase (Enzymatics, Beverly, MA, USA) with 1× blue buffer and 10 mM dATP. The pTRV2-LIC vector was digested with PstI and treated with T4 DNA polymerase in 1× buffer and 10 mM dTTP. T4 DNA polymerase-treated mixtures were incubated at 22°C for 30 min, followed by 75°C for 20 min. The two reaction products were then mixed in a 1:3 (vector:insert) ratio. For ligation, the mixture was incubated at room temperature for 30 min, then transformed into Trans5α competent cell (TransGen Biotech, Beijing, China). Plasmids were extracted using AccuPrep® Plasmid Mini Extraction Kit (Bioneer) and sequenced (Macrogen). Plasmids with complete sequences were introduced into Agrobacterium (Agrobacterium tumefaciens) strain GV3101 by electroporation. pTRV2::PDS and pTRV2::GFP were kindly provided by Prof. Doil Choi (Seoul National University). Primers used for the VIGS study are listed in Table S1.

Agrobacterium infiltration
Agrobacterium carrying pTRV1, pTRV2::PDS, pTRV2::GFP, and pTRV2::CaAN3 were
grown at 28°C for 2 d. Agrobacterium overnight cultures (5 mL) were pelleted by centrifugation
and resuspended in 10 mM 2-(N-morpholino) ethanesulfonic acid buffer (pH 6.0), 10 mM
MgCl$_2$, and 200 µM acetosyringone, at a final OD at 600 nm of 0.6. pTRV1 and pTRV2
constructs were mixed in a 1:1 ratio. After the cell suspensions were incubated at room
temperature for 3 h, the constructs were infiltrated into the first and second foliage leaves. The
infiltrated plants were then grown in a chamber at 23°C with a 16-h light/8-h dark photoperiod.

Transient overexpression in *Nicotiana benthamiana* leaves

The coding sequences for CaAN3 and GFP (as control) were cloned into pCAMBIA2300-
LIC vector harboring the cauliflower mosaic virus 35S promoter. The constructs were
transformed into Agrobacterium strain GV3101 by electroporation. Agrobacterium infiltrations
were performed as above. The cell mixture was infiltrated into young *N. benthamiana* leaves
and anthocyanin accumulation was determined at 7–10 d after infiltration. Primers used for
vector construction are listed in Table S1.

Total anthocyanin quantification

Anthocyanins in *N. benthamiana* leaves were extracted and quantified as previously
described (Mazzucato et al. 2013) with slight modifications. Infiltrated leaf sections were
collected and weighted to 0.5 g each, then transferred into a tube containing 5 mL anthocyanin
extraction solution (1-propanol:HCl:distilled water, 1:1:81, v:v:v). The tubes were boiled in a
water bath for 6 min and then incubated overnight in the dark at room temperature. Absorbance
was recorded spectrophotometrically at 535 nm and 650 nm with a NanoDrop1000 instrument
(NanoDrop Technologies, Wilmington, DE, USA). Total anthocyanin contents were quantified
as the difference in absorbance at 535 nm and 650 nm based and normalized to the fresh weight.
of each sample in grams.

**High-performance liquid chromatography (HPLC) analysis of anthocyanidins**

To measure anthocyanin contents, at least three immature fruits were harvested from each individual plant with two replicates and freeze-dried. Extracts were prepared from 0.1 g of immature fruit samples. The HPLC analysis was performed using an UltiMate 3000RS liquid chromatography system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a reversed-phase Waters-XTerra C18 column (Waters Corporation, Milford, MA, USA). Mobile phase A used 0.5% trifluoro-acetic acid, and phase B used 100% acetonitrile. The separation conditions for phase B were 0–2 min (14%), 2–4 min (17%), 4–7 min (28%), 7–10 min (36%), 10–12 min (60%), and 12–15 min (0%) with a flow rate of 0.6 mL/min. As standards for quantification, cyanidin chloride, delphinidin chloride, malvidin chloride, pelargonidin chloride, and peonidin chloride (Extrasynthese, Genay, France) were used, and total anthocyanin contents were quantified by measuring the absorbance at 520 nm. HPLC was performed at Chungnam National University (Daejeon, Korea).
Results

Segregation of the purple phenotype in the mapping population

To map the CaAN3 locus, we generated an F2 population derived from self-pollinated plants from the C. annuum hybrid cultivar Salad Piment Purple, which shows the same purple immature fruit coloration as MAB2. The F2 population used for mapping consisted of 243 individuals and exhibited a segregation ratio of 2.12:1 (immature purple:immature green), with a $\chi^2$ value of 0.011. This segregation pattern was similar to that seen in the F2 population derived from a cross between the cultivars MAB1 and MAB2. All individuals in the mapping population produced mature red fruits (Fig. 1b).

We genotyped CaAN2 in the F2 mapping population, MAB1, and MAB2 with a previously developed SCAR marker, which distinguishes CaAN2 alleles based on structural variation in the CaAN2 promoter region. MAB2 carries a non-functional CaAN2 allele but is characterized by fruit-specific anthocyanin biosynthesis. Screening the F2 population from Salad Piment Purple showed that all individuals harbor a non-functional CaAN2 allele, as do the MAB1 and MAB2 cultivars (Fig. 1c), indicating that the segregating phenotype is independent from the genotype at CaAN2.

The CaAN3 locus maps to chromosome 10

To identify CaAN3, we combined RNA-seq with BSA. We prepared three RNA pools from six individuals for each phenotypic bulk (18 purple and 18 green plants) and sequenced the resulting libraries on the Illumina platform. We obtained an average of 121,311,533 reads per RNA pool by BSR-seq, reaching a total of 33.6 Gb for the purple pools and 39.9 Gb for the green pools and a coverage of 748× (purple samples) and 890× (green samples). An average of 340 million raw reads mapped to the reference genome Dempsey v1.0, covering
about 96% of the predicted coding regions in the genome (Table 1). We identified 311,679 raw SNPs, of which 63,316 remained after filtering out non-polymorphic and low-quality sites. Using these filtered SNPs, we calculated the Δ(SNP-index) values as the difference between the two different pools.

To delimit the CaAN3 candidate interval, we selected 6,672 SNPs with Δ(SNP-index) values higher than the 99% confidence level, leading to identification of five candidate regions on chromosomes 1, 8, and 10. We observed the highest peak of Δ(SNP-index) values on chromosome 10 corresponding to two candidate regions, 0.15–108.8 Mb and 121.7–239.8 Mb (Fig. 2a).

From the BSR-seq analysis, we estimated the physical position of CaAN3 to be within either the interval 0.15–108.8 Mb or the interval 121.7–239.8 Mb on chromosome 10. We genotyped F2 individuals from the mapping population with HRM-10-63 at 62.9 Mb and SWPm_00416 located at 168.1 Mb, which revealed recombinants and thus indicated that the candidate region likely lies downstream of 168.1 Mb on chromosome 10. We used an additional five markers, developed based on the SNP information obtained from BSR-seq, to refine the map position of the CaAN3 locus, leading to a final candidate region spanning the interval 184.6–186.4 Mb on chromosome 10 of the Dempsey v1.0 reference genome (Fig. 2b).

Dem.v1.00043895 is a candidate gene for CaAN3

Since the Δ(SNP-index) approach did not reveal an obvious candidate for CaAN3 on chromosome 10, we looked for differentially expressed genes (DEGs) between purple and green pools and identified 2,175 significant DEGs. Gene Ontology (GO) term enrichment analysis showed that these DEGs are enriched in twelve GO terms, including secondary metabolic process and carbohydrate metabolic process in the ‘biological process’ category;
transcription factor activity and catalytic activity in ‘molecular function’; and extracellular region in ‘cellular component’ (Fig. S1a). In addition, we detected enrichment for eight Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, four of which were related to anthocyanin biosynthesis such as ‘biosynthesis of secondary metabolites’, ‘metabolic pathways’, ‘phenylpropanoid biosynthesis’, and ‘flavonoid biosynthesis’ (Fig. S1b). The EBGs PAL, C4H, 4CL, CHS, CHI, and F3H did not appear to be differentially expressed between purple and green fruits. By contrast, we detected significant differences in the expression levels of the LBGs DFR, ANS, and 3GT, between purple and green fruits (Fig. S2).

The final candidate region contained 15 annotated genes, of which 8 were not expressed in either purple or green fruits. Three genes, Dem.v1.00043888, Dem.v1.0043915, and Dem.v1.0043916, showed minimal expression levels ranging from 0.01 to 0.41 Log_2(average FPKM + 1). Another three genes, Dem.v1.00043885, Dem.v1.0043901, and Dem.v1.0043918, exhibited relatively high expression levels but were not differentially expressed between purple and green immature fruits. The final gene, Dem.v1.00043895, which was annotated as encoding a MYB-type transcription factor, was not expressed in green fruits but was highly expressed in purple fruits (Table 2), making Dem.v1.00043895 the strongest candidate gene for CaAN3.

The Dem.v1.00043895 promoter region exhibits structural variation

To explore possible sequence variation at the Dem.v1.00043895 locus between purple and green genotypes, we randomly selected individuals from the Salad Piment Purple F_2 population and sequenced the Dem.v1.00043895 promoter and coding regions. As controls, we also sequenced Dem.v1.00043895 from genomic DNA extracted from the leaves of F_2 individuals homozygous for linked markers as well as the MAB2 (purple immature fruit) and MAB1 (green immature fruit) cultivars. We determined that the promoter region is identical in F_2 individuals with the purple allele and MAB2. Likewise, F_2 individuals homozygous for the green allele and
MAB1 had identical promoter structures.

We detected several structural variants in the promoter region between purple and green *CaAN3* alleles: a 4-bp insertion at nucleotide \(-1,428\) (+1 being the ATG), a 7-bp insertion at nucleotide \(-1,069\), a 35-bp insertion at nucleotide \(-887\), and a 15-bp deletion at nucleotide \(-157\) in the promoter region of the green *CaAN3* allele (Fig. 3a). We also identified an A-to-D change between the purple and green predicted proteins at amino acid residue 43. We noticed another amino acid change (at residue 211), but the genotype did not correlate with fruit phenotype (Fig. 3b). Based on these polymorphisms in relation to the measured expression pattern, we hypothesized that variation in the promoter region might be responsible for the loss of *Dem.v1.00043895* expression in green fruits, while the purple *CaAN3* allele is functional and expressed in purple immature fruits.

*Dem.v1.00043895 is specifically expressed in immature fruits*

To investigate the expression pattern of the candidate gene *Dem.v1.00043895*, we collected MAB2 fruits over the course of their development (Fig. S3a) and performed RT-qPCR. The candidate gene was most highly expressed when purple pigments accumulate throughout the fruit, in the very early stages of fruit development. As the fruit matured and the purple pigmentation disappeared, the expression levels of the gene gradually decreased. After 40 d post anthesis (DPA), when fruits lost most of their purple pigmentation and took on their mature red color, the gene showed almost no detectable expression (Fig. S3b). We measured anthocyanidin concentrations by HPLC analysis over the same developmental time course, which revealed a similar pattern that correlates with pericarp color. Anthocyanidin contents were highest in the earliest stages of fruit growth before gradually decreasing as fruits developed. Of the five anthocyanidins (cyanidin, delphinidin, malvidin, pelargonidin, and peonidin) detected by HPLC analysis, delphinidin represented over 99% of the total
anthocyanidin contents (Fig. S3c). Dem.v1.00043895 was specifically expressed in immature fruits at 20 DPA but was expressed to much lower levels in leaves, stems, and flowers, as expected for the CaAN3 candidate gene (Fig. S3d).

Silencing Dem.v1.00043895 blocks purple pigmentation in pericarp

To determine whether Dem.v1.00043895 is CaAN3, we silenced the gene via TRV-mediated VIGS. To this end, we cloned a specific fragment from the Dem.v1.00043895 coding sequence into the silencing vector TRV2. Agrobacterium cultures carrying pTRV2::GFP and pTRV2::PDS were used as negative and positive controls, respectively. We infiltrated leaves from the C. annuum cultivar MAB2 with purple immature fruits; infiltration with pTRV2::GFP did not alter fruit pigmentation, while plants infiltrated with pTRV2::PDS developed bleaching in leaves, indicating that Phytoene desaturase (PDS) is properly silenced. Plants infiltrated with pTRV2::GFP or pTRV::PDS produced fully purple fruits that are undistinguishable from those of the parental cultivar MAB2. In sharp contrast, plants infiltrated with the TRV2 construct targeting Dem.v1.00043895 showed alterations in fruit pigmentation. Although the phenotype of infiltrated plants showed slight differences, most infiltrated plants were characterized by fruits lacking accumulation of anthocyanins. In some plants, fruits remained green throughout their immature stages. Other plants showed pericarp with purple and green sectors with clearly defined margins between the two colors (Fig. 4a). The green sectors of the pericarp never turned purple as fruit development progressed. HPLC analysis from several immature and mature fruits from different plants confirmed that delphinidins do not accumulate in the regions devoid of purple pigmentation (Fig. 4b).

We investigated expression levels of anthocyanin biosynthesis genes in control and silenced fruits: when Dem.v1.00043895 was silenced, we observed a downregulation of transcript levels for several genes. Fruits from plants infiltrated with pTRV::GFP were used as
negative control. Transcripts of both EBGs and LBGs accumulated to lower levels in Dem.v1.00043895-silenced plants. However, the EBGs 4CL and F3H were expressed to comparable levels in Dem.v1.00043895-silenced and control fruits (Fig. 4c). LBGs showed clearer differences, with the downregulation of DFR, F3’5’H, and ANS in Dem.v1.00043895-silenced plants. In particular, DFR and F3’5’H transcripts were barely detectable in Dem.v1.00043895-silenced fruits (Fig. 4c). These results suggested that Dem.v1.00043895 may directly regulate LBG expression in the anthocyanin biosynthetic pathway.

Overexpression of Dem.v1.00043895 induces the accumulation of anthocyanin in N. benthamiana leaves

To further explore the role of Dem.v1.00043895 in anthocyanin biosynthesis, we transiently overexpressed the gene in N. benthamiana leaves by cloning the Dem.v1.00043895 coding sequence in the pCAMBIA2300 vector under the control of the 35S promoter (Fig. S4a). We also infiltrated N. benthamiana leaves with Agrobacterium harboring the pCAMBIA2300:GFP construct as a negative control. About 10 d after infiltration, we noticed an alteration of color around the infiltration sites when overexpressing Dem.v1.00043895, while leaves infiltrated with pCAMBIA2300:GFP showed no change in color (Fig. S4b). We confirmed this visual observation by quantifying anthocyanin accumulation (Mazzucato et al. 2013): extracts from leaves overexpressing Dem.v1.00043895 were purple with anthocyanin contents of 7.4–25.6 per g fresh weight, based on the difference in absorbance at 535 nm and at 650 nm, while extracts from non-infiltrated and negative control plants were slightly green with anthocyanin contents of 1.3–2.2 per g fresh weight (Fig. S4c).

The CaAN2 and CaAN3 expression patterns correlate with fruit and flower color in pepper
We hypothesized that Dem.v1.00043895 is a master regulator of fruit-specific anthocyanin biosynthesis and that structural variation in its promoter region determines the functionality of the gene. To test this possibility, we genotyped and phenotyped several pepper accessions with purple or green immature fruit, alongside MAB1 and MAB2 as controls. We classified the 24 different pepper accessions characterized here into four distinct phenotypic groups: Group I with small purple fruits and purple flowers (eight accessions); Group II with purple fruits and white flowers (fruit-specific pigmentation; four accessions, including MAB2); Group III with pale purple fruits together with a yellow background and white flowers (four accessions); and Group IV with green immature fruits and white flowers (eight accessions, including MAB1) (Fig. S5). Detailed phenotypes of the 24 pepper accessions including leaf and stem color are summarized in Table S2.

We also determined the expression levels of the candidate gene for CaAN3 (Dem.v1.00043895) and CaAN2 in these 24 accessions, revealing several trends. First, CaAN2 expression levels showed a positive correlation with Group-I accessions with purple fruit and flowers. Second, CaAN3 was only expressed in Group-II accessions with purple fruit and white flowers, in agreement with Dem.v1.00043895 being CaAN3. The four accessions from Group III with yellow immature skin color with pale purple pigmentation showed little CaAN2 expression and almost no CaAN3 expression. Finally, Group-IV accessions with green immature fruit had almost no detectable expression of either CaAN2 or CaAN3 (Fig. 5a,b). CaAN2 was expressed only in pepper accessions with purple flowers. Three of these accessions also showed purple pigmentation in leaves (Table S2).

**Structural variation in the promoter region is not directly related to CaAN3 expression**

We developed a SCAR marker set targeting the deletion found in the promoter region (Table S3) to genotype the 24 accessions phenotyped above. The MAB1 accession, which
carries the same green CaAN3 allele as Salad Piment Purple, showed the expected 524-bp amplicon. However, most of the other lines yielded a 1,188-bp PCR product (the same size as in MAB2) regardless of their fruit or flower pigmentation (Fig. S6). The CaAN3 promoter region was identical between the two pepper accessions ‘IT218962’ and ‘KC00134’ (with purple pigmentation for both fruit and flowers) and MAB2 (data not shown), although MAB2 only showed expression of CaAN2. The IT158637 and IT229203 accessions with pale purple pigmentation in fruits produced relatively weak or no amplification despite several attempts.
Discussion

In this study, we identified CaAN3 as encoding an R2R3 MYB TF that regulates fruit-specific anthocyanin accumulation in pepper. We propose that the likely candidate gene for CaAN3 is Dem.v1.00043895 in the Dempsey v1.0 reference genome and is located at 185.2 Mb on chromosome 10. Plants derived from the F2 mapping population and homozygous for either the purple or green CaAN3 allele exhibited distinct structural variants in the CaAN3 promoter region. Furthermore, we confirmed that silencing CaAN3 impairs the biosynthesis of fruit-specific anthocyanins in pepper, while overexpression of CaAN3 was sufficient to induce the accumulation of anthocyanins in N. benthamiana leaves.

R2R3 MYB TFs are well-known regulatory factors that activate the transcription of structural genes in the anthocyanin biosynthesis pathway in various plants (Jaakola 2013). Several R2R3 MYB TFs act as anthocyanin activators in Solanaceous plants such as tomato, pepper, eggplant, and potato (Liu et al. 2018). However, relatively few studies have been conducted in pepper, in which the R2R3 MYB TF CaAN2 positively regulates anthocyanin accumulation in multiple tissues including immature fruits, flowers and leaves (Borovsky et al. 2004; Jung et al. 2019). The insertion of a non-LTR retrotransposon into the CaAN2 promoter region determines the functionality of the CaAN2 allele. The genes F3H, DFR, ANS, and Anthocyanidin 3-O-glucosyltransferase (UFGT) were expressed to lower levels in plants with non-functional CaAN2 alleles, suggesting that CaAN2 regulates several LBGs from the anthocyanin biosynthesis pathway (Jung et al. 2019). However, additional genetic factor(s) likely contribute to the regulation of anthocyanin biosynthesis besides CaAN2, especially in accessions with fruit-specific anthocyanin accumulation, as several pepper have purple fruits despite carrying a non-functional CaAN2 allele.

We undertook a BSR-seq analysis to identify the novel locus CaAN3 that regulates fruit-specific anthocyanin biosynthesis by comparing expression levels between plants with purple
or green fruits at the breaker stage and developed molecular markers to narrow down the candidate region encompassing CaAN3, which affected the expression of several LBGs. Sequence analysis of the strongest candidate locus revealed several InDel polymorphisms in the promoter of the presumed green allele that might be associated with the loss of CaAN3 functionality.

We performed VIGS in pepper and transient overexpression in N. benthamiana leaves to validate the role of the candidate gene Dem.v1.00043895 in regulating the expression of anthocyanin biosynthesis genes. Fruits from silenced plants lost purple pigments at the immature stage and even produced some green immature fruits similar to those seen in other pepper accessions with green fruit. Anthocyanin concentrations in pericarps decreased correspondingly in the silenced fruits. Conversely, overexpression of the candidate gene caused purple pigmentation of N. benthamiana leaves, confirming that anthocyanins accumulate upon overexpression of CaAN3.

We developed a SCAR marker set based on the promoter variants detected between plants with green and purple fruits to genotype 24 pepper accessions. Most accessions, including those producing green immature fruits, exhibited the same genotype as the accession MAB2 with purple fruit and thus carrying the functional allele, contradictory to our hypothesis, although we did observe a positive correlation between expression of the CaAN3 candidate gene and anthocyanin accumulation in immature fruit. In particular, the promoter regions from the KC00134 and IT218962 accessions, which express CaAN2 but not CaAN3, were identical to that of MAB2 with purple immature fruit. Therefore, we conclude that there is no direct correlation between the fruit-specific accumulation of anthocyanins and the structural variation in the CaAN3 promoter region.

Although this study successfully identified a gene that can exert the function expected for CaAN3, further studies will be needed to better understand how it is regulated, such as an
investigation of the underlying transcriptional network. Based on the examples of
transcriptional networks in tomato, we hypothesize that another genetic factor may be involved
in CaAN3-mediated anthocyanin biosynthesis. In tomato, four genes encoding R2R3 MYB TFs
(SlAN2, SlANT1, SlAN1, and SlAN2-like) are clustered on chromosome 10. SlAN2 is mainly
expressed in stems, leaves, and flowers rather than immature fruits, whereas SlAN2-like is
specifically and strongly expressed in immature fruits (Kiferle et al. 2015; Sun et al. 2019).
However, overexpression of either SlAN2 or SlAN2-like is sufficient to induce anthocyanin
accumulation in several tissues such as pericarp and stamens. Phylogenic analysis showed that
SlAN2 regulating anthocyanin mainly in vegetative tissues clusters with the fruit-specific
regulator CaAN3, while the fruit-specific regulator SlAN2-like groups with CaAN2 (Fig. 6).
The functional differences between these TFs from tomato and pepper may therefore have
occurred following the divergence of these two members of the Solanaceae (Quattrocchio et al.
1998).
Unfortunately, correlation or hierarchy between the two tomato R2R3 MYB TFs has not
been studied in detail, although their individual functions are well characterized. The
anthocyanin biosynthesis repressor SiMYBATV, a R3 MYB TF, competes with SlAN2-like
for interaction with SlAN1, a bHLH TF. It will be necessary to test whether a homologous
repressor exists in pepper with a similar function and how the gene affects the expression of
CaAN2 or CaAN3. A genetic approach may also be undertaken by crossing two pepper
accessions: one with functional CaAN2 and CaAN3 alleles but that only expresses CaAN2, and
another accession with a non-functional CaAN2 allele and a functional CaAN3 allele. The
expression of CaAN2 and CaAN3 in the resulting F1 and F2 progeny will inform on the genetic
architecture of anthocyanin accumulation.
In conclusion, this study identified a new regulatory factor involved in anthocyanin
biosynthesis in pepper and provided a hypothesis for the underlying regulatory network. Our
results add to the growing body of work aimed at understanding the anthocyanin transcriptional
network in Solanaceous crops including pepper, with the goal of allowing the engineering of a
pepper cultivar with high contents of anthocyanins beneficial to human health.
Supplementary Information The online version contains supplementary material available at xxx.

Funding This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Technology Commercialization Support Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (821012-03). This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ015881)" Rural Development Administration, Republic of Korea.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Authors’ Contributions JB participated in population generation, phenotype analysis, development of markers, gene functional studies and manuscript writing. NL and TGK participated in the fine-mapping of the gene. JHL participated in the BSR-seq analysis. SYJ participated in initial experiment. BCK participated in the design of the study and editing of the manuscript.

Data Availability Transcriptome analysis of immature fruit color in Capsicum annuum was deposited into Sequence read archive (SRA) database under the accession numbers of SAMN25289192 and SAMN25289193, respectively.

Acknowledgements This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Technology Commercialization Support Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (821012-03). This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ015881)" Rural Development Administration, Republic of Korea.
Commercialization Support Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (821012-03). This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ015881)" Rural Development Administration, Republic of Korea.
References


Ohno S, Ueno M, Doi M (2020) Differences in the CaMYBA genome between anthocyanin-pigmented cultivars and non-pigmented cultivars in pepper (Capsicum annuum). Hort J 89:30–36


network promotes anthocyanin biosynthesis in tomato flesh. Mol Plant 13:42–58


**Table 1.** Summary of BSR-seq analysis of the purple and green bulk samples.

<table>
<thead>
<tr>
<th>Information</th>
<th>Purple</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals in pools</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Number of raw reads</td>
<td>332,378,020</td>
<td>395,491,780</td>
</tr>
<tr>
<td>Total size (Gb)</td>
<td>33.6</td>
<td>39.9</td>
</tr>
<tr>
<td>Coverage</td>
<td>~ 748×</td>
<td>~ 890×</td>
</tr>
<tr>
<td>Number of mapped reads</td>
<td>310,907,661</td>
<td>369,932,886</td>
</tr>
<tr>
<td>Percentage of mapped reads</td>
<td>96.1</td>
<td>96.1</td>
</tr>
</tbody>
</table>
Table 2. Annotated genes in the CaAN3 candidate interval on chromosome 10 and their relative gene expression levels.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Physical location (Mb)</th>
<th>Annotation</th>
<th>Relative gene expression</th>
<th>Green</th>
<th>Purple</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM.v1.00043883 184.6</td>
<td>Similar to MYB1: Transcription factor MYB1 (Actinidia chinensis var. chinensis)</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043884 184.7</td>
<td>Protein of unknown function</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043885 184.8</td>
<td>Heavy metal-associated isoprenylated plant protein 6 (Arabidopsis thaliana)</td>
<td>6.72 6.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043888 185.0</td>
<td>Protein of unknown function</td>
<td>0.00 0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043895 185.2</td>
<td>Similar to MYB1: Transcription factor MYB1 (Actinidia chinensis var. chinensis)</td>
<td>0.00 6.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043899 185.6</td>
<td>Similar to MYB113: Transcription factor MYB113 (Arabidopsis thaliana)</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043901 185.7</td>
<td>Similar to ABC1K1: Protein ACTIVITY of BC1 complex kinase (Arabidopsis thaliana)</td>
<td>4.33 4.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043902 185.9</td>
<td>Similar to ALS3: Protein ALUMINUM SENSITIVE 3 (Arabidopsis thaliana)</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043903 186.0</td>
<td>Similar to HD1: Homeobox protein HD1 (Brassica napus)</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043908 186.1</td>
<td>Protein of unknown function</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043909 186.1</td>
<td>Protein of unknown function</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043910 186.1</td>
<td>Similar to MYB1: Transcription factor MYB1 (Actinidia chinensis var. chinensis)</td>
<td>0.00 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043915 186.2</td>
<td>Protein of unknown function</td>
<td>0.00 0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043916 186.2</td>
<td>Similar to NAC078: NAC domain-containing protein 78 (Arabidopsis thaliana)</td>
<td>0.01 0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEM.v1.00043918 186.4</td>
<td>Similar to UGT86A1: UDP-glycosyltransferase 86A1 (Arabidopsis thaliana)</td>
<td>9.53 10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>Primer</td>
<td>Sequence (5’–3’)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HRM-10-63</strong></td>
<td>F</td>
<td>CAGGAATTTTTAGATTTTACGTCAGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ACCGTGACATATTCCCTCCTGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SWPm_00416</strong></td>
<td>F</td>
<td>GCAAGTTCTAATGATGGTGTTTGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AGATCGGAATAATTCCACGATTAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-43883</strong></td>
<td>F</td>
<td>GTGACAGACACGCACCTAGTTCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CATAATCGGCGATCAACAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-43895B</strong></td>
<td>F</td>
<td>GCAGACTTCGTTTGTTGTTAAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CCACATGTGGAGCCGTAAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-43895C</strong></td>
<td>F</td>
<td>CTTCTACTGCGTACGCTATAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GTGCACTTCCTCTCACCCCT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-185.7</strong></td>
<td>F</td>
<td>AGGCCATTTTAGGGTTCCTTCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ATTGTTCCCTCTGAACGCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-186.4</strong></td>
<td>F</td>
<td>AGTCGATGTATGGAAAAATCGGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GCCACTGACATCTCTAATTTGTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A3-DL-187.1</strong></td>
<td>F</td>
<td>TTTGTAAGAAAATAGAATAGGAGGTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GAGGACCATTATCATATAATACACAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PAL</strong></td>
<td>F</td>
<td>ATTGATTGTGCAAGAAATCAATTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GCTCCACTTTGAGCCACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C4H</strong></td>
<td>F</td>
<td>GATTCCTTCCATTCGGTGTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CCTTCTCCCTTGTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4CL</strong></td>
<td>F</td>
<td>CTGGACAGTGTCTGCTGCAAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GGTTACGGGGCAAAGAAGAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CHS</strong></td>
<td>F</td>
<td>GTGGAACCGTTATCGACCTAGCIA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GTATCAGCTTTGACAGGAAAAAGTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F3H</strong></td>
<td>F</td>
<td>AGCTGTATCATCAAGCAGT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CTGTTCGGCAGACCTCTTCAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DFR</strong></td>
<td>F</td>
<td>AATCGCTCGACATTTCTGCTCATC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CTAACACAGGGAAAGGGCTGTTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ANS</strong></td>
<td>F</td>
<td>CAAATGCCCACACCCAGAATAGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CGCAGTTTGCAGTTACCACCTCTTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CaAN2</strong></td>
<td>F</td>
<td>CATAGCTTTCTACGGCAGACAGATG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TGGACATTGAGAGAAGGTCGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CaAN3</strong></td>
<td>F</td>
<td>AAGGAAAGGGCATGAGGAATG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CAGCTTTGACATTTCTGCTGCTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTIN</strong></td>
<td>F</td>
<td>AGCACTTGGAGATAGGAGAAGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGGAGCAGCACCCGCGTGAATAGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>43895-VIGS</strong></td>
<td>F</td>
<td>CGACGACAAGACCCCTCTCCCTGAGAAGGTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GAGAGGAAAGACCTCCCTAAAAATCCCTACCCCAACCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>43895-OE</strong></td>
<td>F</td>
<td>CGACGACAAGACCCCTCTGAGAAGGTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GAGAGGAAAGACCTCCCTAAAAATCCCTACCCCAACCA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table S2.** Phenotypes of 24 pepper accessions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Code (Fig. S6)</th>
<th>Species</th>
<th>Accession</th>
<th>Immature fruit</th>
<th>Mature fruit</th>
<th>Flower</th>
<th>Leaf</th>
<th>Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>a C. frutescens</td>
<td>IT218962</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>b C. annuum</td>
<td>KC00134</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>c C. annuum</td>
<td>AC08-181</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>P</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>d C. annuum</td>
<td>IT223782</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>e C. annuum</td>
<td>IT286136</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>f C. annuum</td>
<td>IT218615</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>g C. frutescens</td>
<td>IT301140</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>h C. annuum</td>
<td>IT158844</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>i C. annuum</td>
<td>MAB2</td>
<td>P</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>j C. annuum</td>
<td>Mavis</td>
<td>P</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>k C. annuum</td>
<td>Mavras</td>
<td>P</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>l C. annuum</td>
<td>IT305471</td>
<td>P</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>III</td>
<td>m C. annuum</td>
<td>IT286162</td>
<td>P'</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>n C. chinense</td>
<td>IT158637</td>
<td>P'</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>o C. chinense</td>
<td>IT229203</td>
<td>P'</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>p C. annuum</td>
<td>IT218898</td>
<td>P'</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>IV</td>
<td>q C. annuum</td>
<td>KC01237</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>r C. annuum</td>
<td>IT218898</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>s C. annuum</td>
<td>Chilbok No.2</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>t C. annuum</td>
<td>Jeju</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>u C. annuum</td>
<td>IT238052</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>v C. annuum</td>
<td>Yuwolcho</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>w C. annuum</td>
<td>IT238056</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>x C. annuum</td>
<td>MAB1</td>
<td>G</td>
<td>R</td>
<td>W</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

* P, Purple; P’, Pale purple; G, Green; R, Red; W, White
**Table S3.** Primer sequences for the SCAR marker specific for *CaAN3.*

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN3_IDEL_F</td>
<td>GCGGATCAGATTCACTTTCCATCTCTAG</td>
</tr>
<tr>
<td>AN3_INS_R</td>
<td>ATTGAAAAACACATACCTTTTCCATAAGTATATTAAATCT</td>
</tr>
<tr>
<td>AN3_DEL_R</td>
<td>GGTATGGTGACTGAATATCTTTCA</td>
</tr>
</tbody>
</table>
Figure legends

**Fig. 1.** Phenotypes of plant materials and *CaAn3* marker analysis. a *C. annuum* lines MAB2 and MAB1. MAB2 has purple immature fruits, while MAB1 has green immature fruits. Both lines have green leaves and stems. b Fruit phenotype of F\textsubscript{2} plants derived from the hybrid *C. annuum* Salad Piment Purple. Pigmentation in immature fruits segregates into purple and green immature fruits. c Genotyping of the mapping population for *CaAN2*. Plants in the mapping population carry the non-functional *CaAN2* allele but still segregate for purple pigmentation. A previously developed SCAR marker set was used for screening *CaAN2*. Two *C. annuum* lines were used as controls: ‘KC00134’ and ‘Chilbok No.2’ for the *CaAN2* functional and non-functional alleles, respectively.

**Fig. 2.** Genetic mapping of *CaAN3* using BSR-seq a Estimation of the candidate interval for *CaAN3*. A total of 6,672 significant SNPs were identified by calculating Δ(SNP-index) between the two BSR-seq datasets. The putative genomic region of *CaAN3* is indicated in yellow b Fine-mapping of *CaAN3*. *CaAN3* was mapped to chromosome 10, between 184,603 kb and 186,432 kb. The number of recombinants from 243 F\textsubscript{2} individuals is indicated below the black bar. The genotypes of F\textsubscript{2} plants homozygous for the green or purple *CaAN3* allele or heterozygous are shown as green, purple, and brown bars, respectively. Seven bars at the bottom represent the genotypes of informative individuals with recombination events.

**Fig. 3.** Structural variation between the purple and green *CaAN3* alleles. a Major sequence variation between the *CaAN3* promoter from putative functional and non-functional alleles.
Purple homozygous individuals of Salad Piment Purple F\textsubscript{2} and MAB2 have identical sequences. Green homozygous individuals of Salad Piment Purple F\textsubscript{2} and MAB1 have identical sequences. Due to the expression pattern of Dem.v1.00043895, the \textit{CaAN3} purple allele is expected to represent a functional allele, while the green \textit{CaAN3} allele is non-functional. b Comparison of the predicted protein sequence between two purple alleles and two green alleles.

**Fig. 4.** Silencing the candidate gene (Dem.v1.00043895) via VIGS. a Representative phenotypes of fruits from the wild type (WT), \textit{VIGS::GFP}, \textit{VIGS::PDS}, and \textit{VIGS::CaAN3} plants. \textit{VIGS::GFP} and \textit{VIGS::PDS} plants were used as negative and positive controls, respectively, and only produced purple fruits. \textit{VIGS::CaAN3} plants exhibited a partial or complete loss of purple pigment accumulation. Blue arrow indicates bleached leaf, indicating that \textit{PDS} is properly silenced. Red arrows indicate loss of purple pigmentation in \textit{CaAN3}-silenced fruits. b HPLC analysis of anthocyanidin concentrations in the pericarp. Anthocyanidins were measured by pooling the green (or red) parts of the pericarps sampled from three different fruits. A similar trend was also observed in mature fruits (pool 4). Most of the anthocyanidins measured were delphinidins. c Relative gene expression in the pericarp of fruits silenced for Dem.v1.00043895. Relative expression levels of four early anthocyanin biosynthetic genes (\textit{C4H, 4CL, CHS, F3H}), three late anthocyanin biosynthetic genes (\textit{F3’5’H, DFR, ANS}), and the \textit{CaAN3} candidate gene in the pericarps of wild-type, \textit{GFP}-silenced fruits (negative control), and Dem.v1.00043895-silenced fruits. Asterisks indicate significant difference (\(p < 0.05\)).
Expression analysis of CaAN2 (a) and CaAN3 (b) in different pepper accessions. Group I (from IT218962 to IT158844) with both purple fruits and flowers showed relatively high CaAN2 expression but almost no CaAN3 expression. In Group II accessions (from MAB2 to IT305471), CaAN2 was barely expressed while CaAN3 was highly expressed. Group III (from IT286162 to AC09-003) showed low CaAN2 expression and almost no CaAN3 expression. Group IV with no purple pigmentation in either fruits or flowers had the lowest expression levels of CaAN2 and CaAN3.

Fig. 5. Expression analysis of CaAN2 (a) and CaAN3 (b) in different pepper accessions.

Phylogenetic analysis of CaAN2, CaAN3, and tomato R2R3 MYB TFs.

Fig. 6. Phylogenetic analysis of CaAN2, CaAN3, and tomato R2R3 MYB TFs.

Transcriptome analysis using fruit-specific DEGs. a GO term enrichment analysis. GO analysis was performed for the three categories biological process, molecular function, and cellular component. b KEGG pathway enrichment analysis, using 2,175 DEGs between green and purple immature fruits. The red box highlights pathways known to be related to anthocyanin biosynthesis.

Fig. S1. Transcriptome analysis using fruit-specific DEGs. a GO term enrichment analysis. GO analysis was performed for the three categories biological process, molecular function, and cellular component. b KEGG pathway enrichment analysis, using 2,175 DEGs between green and purple immature fruits. The red box highlights pathways known to be related to anthocyanin biosynthesis.

Expression of anthocyanin biosynthetic genes in the pericarp of green and purple pools from the F₂ population derived from the hybrid Salad Piment Purple. a Core anthocyanin biosynthesis pathway in pepper. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl CoA; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonoid 3-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; 3GT, flavonoid 3-O-glucosyltransferase. b Expression levels of structural genes in the
anthocyanin biosynthetic pathway. The numbers indicate gene expression levels as \( \log_2(\text{mean FPKM} + 1) \), of each pool.

**Fig. S3.** Physiological changes of MAB2 fruits during development. **a** Phenotypic changes of fruit pigmentation from 10 to 50 DPA. **b** Relative Dem.v1.00043895 expression levels as a function of fruit development. **c** Changes in concentration of delphinidins in the pericarp. Most of the anthocyanidins measured are delphinidins. **d** Relative Dem.v1.00043895 expression levels in different tissues. Dem.v1.00043895 showed higher expression levels immature fruit compared to leaves, stems, and flowers at 20 DPA.

**Fig. S4.** Transient overexpression of CaAN3 in *N. benthamiana* leaves. **a** Schematic diagram of the CaAN3 effector construct, consisting of the CaAN3 coding sequence placed under the control of the cauliflower mosaic virus 35S promoter. **b** Phenotypic alterations caused by CaAN3 overexpression in *N. benthamiana* leaves. *GFP* was overexpressed as a negative control. **c** Anthocyanin extracts and total anthocyanin contents for each extract measured by spectrophotometry. Each transiently infiltrated leaf sector with 35S:CaAN3 showed some accumulation of anthocyanins over levels detected in 35S:GFP leaves and control non-infiltrated leaves.

**Fig. S5.** Phenotypes of immature fruits and flowers from 24 pepper accessions. Four phenotypic patterns were observed. **a–h** have both purple fruits and purple pigmented flowers. **i–l** exhibit fruit-specific purple pigmentation. **m–p** display pale purple pigmentation with a yellowish background. **q–x** have no purple pigmentation in either fruits or flowers. **a** IT218962 **b**
Fig. S6. Testing the CaAN3 SCAR marker on 24 pepper accessions. Although there are different expression patterns of CaAN3 among 24 pepper lines, genotypic variation was not correlated. Except for MAB1, most other pepper lines have the identical structure as MAB2 with fruit-specific anthocyanin accumulation. IT229203 might harbor another type of promoter structure, since we failed to amplify a PCR product despite several attempts.
Figure 1

Phenotypes of plant materials and CaAn3 marker analysis. a C. annuum lines MAB2 and MAB1. MAB2 has purple immature fruits, while MAB1 has green immature fruits. Both lines have green leaves and stems. b Fruit phenotype of F2 plants derived from the hybrid C. annuum Salad Piment Purple. Pigmentation in immature fruits segregates into purple and green immature fruits. c Genotyping of the mapping population for CaAN2. Plants in the mapping population carry the non-functional CaAN2 allele but still segregate for purple pigmentation. A previously developed SCAR marker set was used for screening CaAN2. Two C. annuum lines were used as controls: ‘KC00134’ and ‘Chilbok No.2’ for the CaAN2 functional and non-functional alleles, respectively.

Figure 2

Genetic mapping of CaAN3 using BSR-seq a Estimation of the candidate interval for CaAN3. A total of 6,672 significant SNPs were identified by calculating $\Delta$(SNP-index) between the two BSR-seq datasets. The putative genomic region of CaAN3 is indicated in yellow b Fine-mapping of CaAN3. CaAN3 was mapped to chromosome 10, between 184,603 kb and 186,432 kb. The number of recombinants from 243 F2 individuals is indicated below the black bar. The genotypes of F2 plants homozygous for the green or purple CaAN3 allele or heterozygous are shown as green, purple, and brown bars, respectively. Seven bars at the bottom represent the genotypes of informative individuals with recombination events.
Figure 3

Structural variation between the purple and green CaAN3 alleles. a Major sequence variation between the CaAN3 promoter from putative functional and non-functional alleles. Purple homozygous individuals of Salad Piment Purple F2 and MAB2 have identical sequences. Green homozygous individuals of Salad Piment Purple F2 and MAB1 have identical sequences. Due to the expression pattern of Dem.v1.00043895, the CaAN3 purple allele is expected to represent a functional allele, while the green CaAN3 allele is non-functional. b Comparison of the predicted protein sequence between two purple alleles and two green alleles.
Figure 4

Silencing the candidate gene (Dem.v1.00043895) via VIGS. a Representative phenotypes of fruits from the wild type (WT), VIGS::GFP, VIGS::PDS, and VIGS::CaAN3 plants. VIGS::GFP and VIGS::PDS plants were used as negative and positive controls, respectively, and only produced purple fruits. VIGS::CaAN3 plants exhibited a partial or complete loss of purple pigment accumulation. Blue arrow indicates bleached leaf, indicating that PDS is properly silenced. Red arrows indicate loss of purple pigmentation in CaAN3-
silenced fruits. b HPLC analysis of anthocyanidin concentrations in the pericarp. Anthocyanidins were measured by pooling the green (or red) parts of the pericarps sampled from three different fruits. A similar trend was also observed in mature fruits (pool 4). Most of the anthocyanidins measured were delphinidins. c Relative gene expression in the pericarp of fruits silenced for Dem.v1.00043895. Relative expression levels of four early anthocyanin biosynthetic genes (C4H, 4CL, CHS, F3H), three late anthocyanin biosynthetic genes (F3′5′H, DFR, ANS), and the CaAN3 candidate gene in the pericarps of wild-type, GFP-silenced fruits (negative control), and Dem.v1.00043895-silenced fruits. Asterisks indicate significant difference (p < 0.05).

Figure 5

Expression analysis of CaAN2 (a) and CaAN3 (b) in different pepper accessions. Group I (from IT218962 to IT158844) with both purple fruits and flowers showed relatively high CaAN2 expression but almost no CaAN3 expression. In Group II accessions (from MAB2 to IT305471), CaAN2 was barely expressed while CaAN3 was highly expressed. Group III (from IT286162 to AC09-003) showed low CaAN2 expression and almost no CaAN3 expression. Group IV with no purple pigmentation in either fruits or flowers had the lowest expression levels of CaAN2 and CaAN3.

Figure 6

Phylogenetic analysis of CaAN2, CaAN3, and tomato R2R3 MYB TFs.

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

- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf
- Fig.S4.pdf
- Fig.S5.pdf
- Fig.S6.pdf