A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Identification of the *CaAN3* Gene that Regulates Fruit-specific Anthocyanin Biosynthesis in Pepper (*Capsicum annuum*)

고추의 과실특이적 안토시아닌 생합성 조절 유전자 *CaAN3* 동정

FEBRUARY, 2022

JINYOUNG BYUN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF AGRICULTURE, FORESTRY, AND BIORESOURCES
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Identification of the *CaAN3* Gene that Regulates Fruit-specific Anthocyanin Biosynthesis in Pepper (*Capsicum annuum*)

UNDER THE DIRECTION OF DR. BYOUNG-CHEORL KANG
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

BY
JINYOUNG BYUN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF AGRICULTURE, FORESTRY, AND BIORESOURCES
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

FEBRUARY, 2022

APPROVED AS A QUALIFIED DISSERTATION OF
JINYOUNG BYUN
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Hee Jae Lee, Ph.D.

VICE-CHAIRMAN _____
Byoung-Cheorl Kang, Ph.D.

MEMBER _____
Eun Jin Lee, Ph.D.

Identification of the *CaAN3* Gene that Regulates Fruit-specific Anthocyanin Biosynthesis in Pepper (*Capsicum annuum*)

JINYOUNG BYUN

DEPARTMENT OF AGRICULTURE, FORESTRY, AND BIORESOURCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSRACT

The key regulatory gene *CaAN2* encoding R2R3 Myeloblastosis (MYB) transcription factor has been known to regulate anthocyanin biosynthesis in various tissues in pepper. *CaAN2* is not expressed in certain pepper accessions showing fruit-specific anthocyanin accumulation. In this study, a novel locus named *CaAN3* that regulates fruit-specific anthocyanin biosynthesis was identified. An F_2 population segregating for *CaAN3* was constructed by self-pollination of an F_1 hybrid cultivar showing purple immature fruits. Total RNA was extracted from the F_2 plants and separately pooled according to purple and green phenotypes. Three RNA pools containing six individuals were prepared for each phenotype. Bulked segregant RNA sequencing was conducted for two different RNA pools via the Illumina platform.

Raw sequences were aligned to the pepper reference genome ‘Dempsey’, and a total of 6,672 significant single nucleotide polymorphisms (SNPs) were identified by calculating Δ(SNP – index) between the two pools. Molecular mapping was conducted to delimit the target region of *CaAN3* spanning 184.6 - 186.4 Mbp of chromosome 10. Fifteen candidate genes were annotated in the target region, and *Dem.v1.00043895*, an R2R3 MYB transcription factor, was selected from them as the strongest candidate gene based on the differentially expressed genes analysis result. Sequence analysis revealed that there are four indel variations in the promoter region of the *CaAN3* green allele. Virus-induced gene silencing and transient overexpression assay were performed to characterize the function of the candidate gene. When *Dem.v1.00043895* was silenced, anthocyanin accumulation in pericarps was significantly reduced. When *Dem.v1.00043895* was overexpressed in *Nicotiana benthamiana* leaves, anthocyanins accumulated around the inoculation sites. These results showed that *Dem.v1.00043895* functions as *CaAN3*, an activator of anthocyanin biosynthesis in pepper fruits. Gene expression analysis and genotypic screening were performed using 24 different pepper accessions to test whether *CaAN3* functions generally. *CaAN3* was expressed only in the fruit-specific purple accessions. However, genotype analysis revealed that the structural variations in the promoter region were not directly related to the expression of *CaAN3*, indicating that another genetic factor is involved in anthocyanin biosynthesis in pepper fruit.

Key words: Anthocyanin, Bulked segregant RNA sequencing, Fine mapping, Fruit-specific anthocyanin accumulation, Pepper, R2R3 MYB transcription factor, Transient overexpression assay, Virus-induced gene silencing.

Student number: 2020-22248

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AN2	*Anthocyanin 2* (gene)
AN3	*Anthocyanin 3* (gene)
bHLH	Basic helix-loop-helix
BSA	Bulked segregant analysis
BSR-Seq	Bulked segregant RNA sequencing
DEG	Differentially expressed gene
DPA	Days post anthesis
EBG	Early biosynthesis gene
GO	Gene ontology
HPLC	High-performance liquid chromatography
HRM	High resolution melting
KEGG	Kyoto Encyclopedia of Genes and Genomes
LBG	Late biosynthesis gene
LIC	Ligation independent cloning
MBW	MYB-bHLH-WD40
MYB	Myeloblastosis
NGS	Next generation sequencing
PCR	Polymerase chain reaction
qRT-PCR	Quantitative reverse transcription PCR
QTL	Quantitative trait locus
SCAR	Sequence-characterized amplified region
SNP	Single nucleotide polymorphism
TF	Transcription factor
VIGS	Virus-induced gene silencing

INTRODUCTION

Anthocyanins are secondary metabolites of plants 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 of plants, primarily red, blue, and purple. Anthocyanins have core functions in plants such as attracting pollinators and protecting plants from biotic and abiotic stressors (de Pascual-Teresa et al., 2010; Jimenez-Garcia et al., 2013). Anthocyanins are gaining popularity due to their favorable impacts on human health. Anthocyanins, in particular, have been shown to have anti-inflammatory and anti-carcinogenic properties in the human body, 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 are critical components that contain not only aesthetic values but also health benefits, breeding cultivars with increased anthocyanin contents has become a priority in various crops (Allan & Espley, 2018).

The anthocyanin biosynthesis pathway has extensively been described in a variety of horticultural crops, as well as model plants *Arabidopsis* and petunia (Jaakola, 2013; Liu et al., 2018; Pelletier et al., 1997; Tsukaya et al., 1991). The anthocyanin biosynthesis pathway, as a branch of the flavonoid biosynthesis pathway, consists of two types of genes: structural genes that encode the enzymes directly participating in the synthesis of anthocyanins, 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).

Coordinate activation of anthocyanin pathway structural genes has been demonstrated, and transcription factors (TFs) that directly regulate the expression of the structural genes have been identified in several plant species (Jaakola, 2013). The interaction of DNA-binding R2R3 myeloblastosis (MYB) TFs with MYC-like basic helix-loop-helix (bHLH) and WD40-repeat proteins regulates anthocyanin biosynthesis (Koes et al., 2005). They generally function in the form of the MBW (MYB-bHLH-WD40) complex, of which the most crucial component is known as the MYB TF, and MYB TF has been reported to activate the 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, eggplant, potato, and pepper. Several crop species of the Solanaceae have been reported to accumulate anthocyanins (Dhar et al., 2015). Due to the commercial importance of Solanaceous crops, substantial researches have been conducted on the anthocyanin regulatory mechanisms (Liu et al., 2018). Among them, anthocyanin synthesis in edible parts such as fruits or tubers has been mainly studied. In tomato, *SlANT1* and *SlAN2* have been identified as R2R3 MYB activators, causing anthocyanin accumulation in vegetative tissues and fruits (Povero et al., 2011;

Schreiber et al., 2012). *SlAN2-like*, another R2R3 MYB TF driven by fruit-specific promoter, was also reported to act as a master regulator in anthocyanin biosynthesis, activating both structural and regulatory genes (Sun et al., 2019). In eggplant, functional R2R3 MYB *SmMYB1* and *SmMYBc* have been reported to upregulate the structural genes and anthocyanin contents in fruits (Gisbert et al., 2016; Stommel & Dumm, 2015; Zhang et al., 2014). In potato, *StAN1*, *StMYBA*$_1$, and *StMYB113* encoding R2R3 MYB TFs were reported as activators of anthocyanin biosynthesis that have high expression levels in purple colored tubers and correlated positively with the transcription of structural genes and with anthocyanin concentration (Liu et al., 2016; Strygina et al., 2019).

In pepper, relatively few studies on regulatory genes have been reported. Among them, *CaAN2*, a R2R3-MYB TF that regulates anthocyanin biosynthesis has been identified (Borovsky et al., 2004). When a non-LTR retrotransposon was inserted into the *CaAN2* promoter region, *CaAN2* was expressed and subsequently, anthocyanins were synthesized. Various tissues including fruits (only at the immature stage) and flowers showed purple pigmentation when *CaAN2* was functional. If the structural genes were not properly transcribed, the immature fruits remained green (Jung et al., 2019; Ohno et al., 2020). However, *CaAN2* was not detected in all instances of the purple pigmentation in pepper fruits, since certain pepper accessions having a non-functional *CaAN2* allele still showed immature fruits with full purple pigmentation (Jung, 2019). These pepper accessions showed purple pigmentation only in their fruits. However, the other accessions having a functional

CaAN2 allele showed purple pigmentation in flowers and leaves as well as in fruits.

Bulked segregant analysis (BSA) can be used to identify markers linked to any specific gene or genomic region using two bulked DNA sample pools. Each bulk contains individuals that have an identical target trait or genomic region but are arbitrary at all unlinked regions (Michelmore et al., 1991). Bulked segregant RNA sequencing (BSR-Seq) method, the modification of BSA that makes use of RNA instead of DNA, makes it possible to efficiently isolate genes (Liu et al., 2012). RNA sequencing is also one of the most widely adopted methods using next generation sequencing technology, which enables comparative quantification of gene expression according to the phenotypic variation (Marioni et al., 2008). Furthermore, RNA sequencing can be applied to investigate structural variations of DNA coding regions such as single nucleotide polymorphisms (SNPs), which can also be used as genetic markers (Chepelev et al., 2009).

In this study, a novel locus *CaAN3* regulating fruit-specific anthocyanin accumulation was identified using BSR-Seq. *CaAN3* has previously been proposed to regulate fruit-specific anthocyanin biosynthesis (Jung, 2019). The previous study developed an F_2 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. The phenotypic segregation ratio of the population was observed to be 3 (purple) to 1 (green), indicating that *CaAN3* is a single dominant locus. Genetic mapping suggested that *CaAN3* was located on chromosome 10, but fine mapping could not be performed

due to the wide range of non-recombination region (Jung, 2019). To fine map the *CaAN3*, a hybrid bell pepper cultivar with the non-functional *CaAN2* allele, which shows purple pigmentation only in fruits was used to develop an F_2 population. The F_2 population derived from this hybrid cultivar showed phenotypic segregation of purple and green immature fruits, enabling the identification of the novel gene. Here, fine mapping of the *CaAN3* locus was performed to identify a candidate gene for *CaAN3*. In addition, virus-induced gene silencing (VIGS) and transient overexpression assay were applied to examine whether the candidate gene actually regulate the fruit-specific anthocyanin accumulation in pepper. Finally, the expression mechanism of the *CaAN3* gene was attempted to be determined using diverse pepper accessions that show purple pigmentation.

MATERIALS AND METHODS

Plant materials

Two *C. annuum* lines 'MAB2' and 'MAB1' were obtained from Asia Seed Co., Ltd. (Incheon, Korea) and used as controls for phenotype and genotype analyses. 'MAB2' has purple immature fruits, but green leaves and stems while 'MAB1' has green immature fruits, leaves, and stems (Fig. 1.). In particular, fruit-specific purple line 'MAB2' was utilized in subsequent experiments after the candidate gene was identified. To map the *CaAN3* locus regulating fruit-specific anthocyanin biosynthesis, *C. annuum* hybrid bell pepper purple cultivar 'Salad Piment Purple' from Takii Seed Co., Ltd. (Kyoto, Japan) was purchased from the seed market to generate an F_2 segregating population. The F_2 population of 243 individuals was used for the bulked RNA-Seq analysis. In addition, thirteen pepper lines obtained from National Institute of Agricultural Science (Wanju, Korea) were used to validate the identity of *CaAN3*.

Nucleic acid extraction

Genomic DNAs were extracted from fresh young leaves of the plants according to the cetyltrimethylammonium bromide protocol (Lee et al., 2017). The extracted gDNA was dissolved in 1× TE buffer and then diluted to a concentration of 50 ng/μL with tri-distilled water. Pericarp tissues of immature fruit were used for the bulked

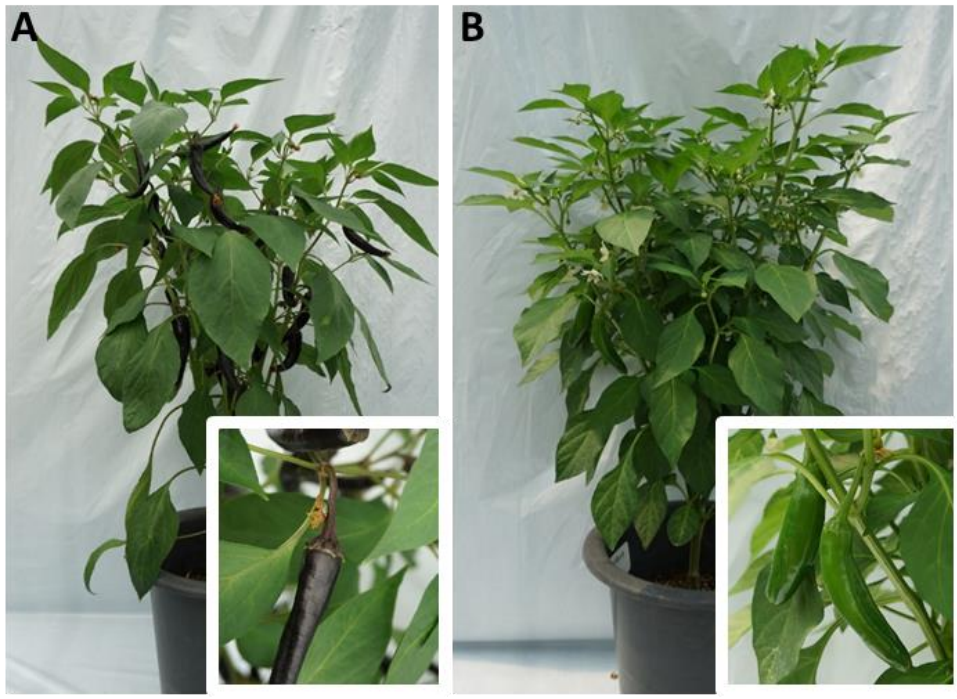

Fig. 1. Phenotypes of *C. annuum* lines 'MAB2'(A) and 'MAB1'(B). 'MAB2' and 'MAB1' have purple and green immature fruits, respectively. Both lines have green leaves and stems.

RNA-Seq analysis and expression analyses. Total RNA was extracted from pericarp at the immature stage of fruit using MG RNAzol Kit (MGmed, Seoul, Korea) according to the manufacturer's instruction.

BSR-Seq

For BSR-Seq analysis, the equal amount of RNA was sampled from eighteen purple and green fruits each from the F_2 population, and pooled to make each of three pools containing six individual RNA samples. TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) was used to construct RNA-Seq library and RNA sequencing was performed in Macrogen (Seoul, Korea). Raw reads of sequences were aligned to pepper reference genome Dempsey v1.0 (unpublished), utilizing STAR version 2.7.5a (Dobin et al., 2013). Then, the alignments of purple and green RNA pools were analyzed following quantitative trait locus (QTL) sequencing analysis pipeline (Takagi et al., 2013). SNPs were extracted from the alignment files of two pools using Samtools. SNPs were filtered and plotted by internal Perl and R scripts of QTL-Seq pipeline. SNP-index value in each pool was computed by the number of aligned reads on the reference genome. Consequently, $\Delta(\mathrm{SNP-index})$ value was defined as difference of $\mathrm{SNP-index}$ value between purple and green RNA pools.

Development of molecular marker based on *CaAN3*

The previously delimited *CaAN3* target region in CM334 reference genome

(Jung, 2019) was used for further marker development. Based on SNPs identified from BSR-Seq method, markers for high resolution melting (HRM) were developed with the amplicon size of 100 to 300 bp, if polymorphisms were detected in polymerase chain reaction (PCR) amplicons. Three primers for the sequence-characterized amplified region (SCAR) marker were developed to amplify the promoter region targeting the insertion and deletion between purple and green alleles. PCR primers were designed by Primer3 software (http://web.bioneer.co.kr/cgibin/primer/primer3.cgi). HRM markers were mainly used for fine mapping of the *CaAN3* locus, and the SCAR marker was used to genotype various different pepper accessions. Primers used as molecular markers are listed in Table 1.

HRM analysis

HRM markers were developed based on the BSR-Seq result to test polymorphisms of developed markers in the F_2 population. HRM markers were used to genotype F_2 individuals using Rotor-Gene 6000 real-time PCR thermocycler (Corbett Research, Sydney, Australia). The real-time PCR was firstly performed with PCR mixture containing 2.5 μL of 10× HiPi reaction buffer (Elpis Biotech, Daejeon, Korea), 2 μL of 10 mM dNTPs, 0.5 μL each of 10 pmol primers, 2 μL of 50 ng/μL gDNA, 0.3 μL Home *Taq* polymerase and sterile distilled H_2O up to 20 μL. The real-time PCR was conducted with 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 step, HRM analysis was carried out with increasing temperature of 0.1 °C every minute from 65 to 95 °C.

Table 1. Primers used in this study

Purpose	Primer	Sequence (5' – 3')
Mapping of *CaAN3*	HRM-10-63_F	CAGGAATTTTTAGATTAGGTTCAGA
	HRM-10-63_R	ACCGTCATATTACCTCCCTGAG
	SWPm_00416_F	GCAAGTTCTAATGATGGTTGTG
	SWPm_00416_R	AGATCCGAAATTTCCACAGTAC
	A3-DL-43883_F	GTCAGACACGCCACTAGTTCA
	A3-DL-43883_R	CATAATCGGCGATCAACAAC
	A3-DL-43895B_F	GCAGACTTCGGTGGTTGAAT
	A3-DL-43895B_R	CCAGCATGTTGAGGCATAAT
	A3-DL-43895C_F	CCTTCACTGCGTACGGCTAT
	A3-DL-43895C_R	GTGCCACTTTCCTTCACCAT
	A3-DL-185.7_F	AGGCAATTTTAGGGTCTCTTTTCA
	A3-DL-185.7_R	ATTGTTCCCTTGAAACGCAGG
	A3-DL-186.4_F	AGTCGATGATTGGAAAATCGGGA
	A3-DL-186.4_R	GCCACTGCATCCTCTAATTCGT
	A3-DL-187.1_F	TGGTGAAGAAAATGAAATGGAGGTC
	A3-DL-187.1_R	TGGACCATTTCACTAAATAACACAA
Gene expression analysis	PAL_F	ATTGATTTTTGCAAGAAATCAATTC
	PAL_R	GCTCCACTTTAGCCCCAC
	C4H_F	GATTCCTTCCATTCGGTGTT
	C4H_R	CCTTTCTCCGTGGTGTCG
	4CL_F	CTGGACCAGTGCTGGCAAT
	4CL_R	GGTTACGGGGCAAAGAACAA
	CHS_F	GTGGAACCGTTATCCGACTAGCAA
	CHS_R	GTATCACTTGGGCCACGGAAAGTA
	F3H_F	ACGCTGATCATCAAGCAGTG
	F3H_R	CTTTTCGGCAACCTCTTCAG
	DFR_F	AATCGCTCCAGCTGGTCTCATCAT
	DFR_R	CTAACACAGGGAAGAGGCTGGTTT
	ANS_F	CAAATGCCCACAACCAGAACTAGC
	ANS_R	CGCACTTTGCAGTTACCCACTTTC
	CaAN2_F	CATAAGCTTCTAGGCAACAGATG
	CaAN2_R	TGCACTTGATGAGAAGGTCCG
	CaAN3_F	AAGGAAAGGTGCATGGAATG
	CaAN3_R	CACCTCTCTTGATATGTGGC
	ACTIN_F	AGCAACTGGGACGATATGGAGAAG
	ACTIN_R	AAGAGACAACACCGCCTGAATAGC
VIGS study	43895-VIGS_F	CGACGACAAGACCCTCTCGCCTTCTGAGGAAGGTA
	43895-VIGS_R	GAGGAGAAGAGCCCTAAAATTCATCCCAACCACCA
Transient overexpression	43895-OE_F	CGACGACAAGACCCTATGAATACTCCAATAATCTGT ACAACATCG
	43895-OE_R	GAGGAGAAGAGCCCTCTAATTAAGTAGATTCCATAT GTCATCCAT

Quantitative reverse transcription (qRT) PCR

Total RNA was seperately extracted from pericarps of the subject plants as described above, and cDNA was synthesized using AccuPower® RT PreMix (Bioneer, Daejeon, Korea) following the manufacturer's protocol. The expression levels of anthocyanin biosynthetic genes including *CaAN3* were measured using Rotor-Gene 6000 real-time PCR thermocycler (Corbett Research). Primers of structural genes were modified from the previous report (Zhang et al., 2015). Real-time PCR was performed using the following conditions: 95 °C for 4 min, followed by 45 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s. The relative expression levels of the genes were normalized against the threshold cycle value of the internal reference gene, *Actin* (Schmittgen & Livak, 2008). Primers used for the gene expression analysis were listed in Table 1.

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. Purified PCR amplicon was treated with T4 DNA polymerase (Enzymatics, Beverly, MA, USA) with 1× blue buffer and 10 mM dATP. *Pst*I digested pTRV2-LIC vector was treated with T4 DNA polymerase with 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 tumefaciens* strain GV3101 through electroporation with 2.0 kV. pTRV2::PDS and pTRV2::GFP were kindly provided by Prof. Doil Choi at Seoul National University. Primers used for VIGS study were listed in Table 1.

Agrobacterium infiltration

Agrobacterium carrying pTRV1, pTRV2::PDS, pTRV2::GFP, and pTRV2::CaAN3 were grown at 28 °C for 2 days. The *Agrobacterium* cultures of 5 mL were resuspended in 10 mM 2-(N-morpholino)ethanesulfonic acid buffer (pH 6.0) including 10 mM $MgCl_2$, and 200 μM acetosyringone to obtain final absorbance at 600 nm of 0.6. pTRV1 and pTRV2 constructs were mixed in an 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 inoculated plants were grown in a chamber at 23 °C with a 16-h photoperiod.

Transient overexpression analysis in *Nicotiana benthamiana* leaves

For the transient transformation expression analysis, two constructs carrying *CaAN3* and *GFP* (as control reaction) were cloned into pCAMBIA2300-LIC vector

containing Caulimosaic virus 35S promoter. The constructs were transformed into *A. tumefaciens* strain GV3101 using an electroporation method, and then transformed cells were incubated at 30 °C for 2 days. Similar method was used to *Agrobacterium* infiltration. The mixture was injected into young *N. benthamiana* leaves and anthocyanin accumulation was determined at 7 to 10 days after infiltrations. Primers used for construction of vector were listed in Table 1.

Total anthocyanin quantification

Anthocyanins in leaves of *N. benthamiana* were extracted and quantified as previously described (Mazzucato et al., 2013) with slight modifications. The sections of leaves with the alteration of pigment were sampled and weighted to 0.5 g each, then transferred into a tube containing 5 mL of anthocyanin extraction solution (1-propanol:HCl:distilled water, 1:1:81, v:v:v). Then the tubes were boiled in a water bath for 6 min. After that, those tubes were incubated overnight in darkness at room temperature. Absorbance at 535 nm and 650 nm were recorded spectrophotometrically with the NanoDrop1000 (NanoDrop Technologies, Wilmington, DE, USA). Total anthocyanins were quantified as the differential of absorbance value of 535 nm and 650 nm based on the gram fresh weight of each leaf samples.

High-performance liquid chromatography (HPLC) analysis of anthocyanidins

To measure the anthocyanidin contents in the pericarps of ‘MAB2’ according to developmental stages and after gene silencing, more than three immature fruits

were harvested from each individual plant with two replicates. Freeze dried 0.1 g of immature fruit samples was used to perform HPLC. The HPLC analysis was performed using 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). 0.5% trifluoroacetic acid was used for mobile phase A, and 100% acetonitrile was used for B. 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 the 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 the total anthocyanidin content was quantified by measuring the absorbance at 520 nm. The HPLC was performed at Chungnam National University (Daejeon, Korea).

RESULTS

Mapping population showed segregation of the purple phenotype

To map the *CaAN3* locus, an F_2 population was constructed using *C. annuum* 'Salad Piment Purple' cultivar that showed the same purple immature fruit phenotype as 'MAB2'. The size of the F_2 population for mapping was 243, and the segregation ratio was 2.12:1 (immature purple:immature green), with 0.011 of χ^2 value. The mapping population showed the same segregation pattern as shown in the F_2 population derived from 'MAB2' and 'MAB1'. All individuals in the mapping population showed mature red fruits (Fig. 2A).

To confirm whether the purple phenotype in the 'Salad Piment Purple' F_2 population is expressed by the same mechanism as 'MAB2', previously developed SCAR marker was used for *CaAn2* genotyping. This marker was developed to distinguish functionality of the regulatory *CaAN2* gene based on the structural variation in the promoter region of *CaAN2*. 'MAB2' was known to have non-functional *CaAN2* but reported to have fruit-specific anthocyanin biosynthesis. Screening for the F_2 population from 'Salad Piment Purple' showed that all individuals contained non-functional *CaAN2* allele, which is identical to 'MAB' population (Fig. 2B). The result indicated that, as in the case of 'MAB2', no regulatory event by *CaAN2* occurred during fruit-specific purple pigmentation for 'Salad Piment Purple'.

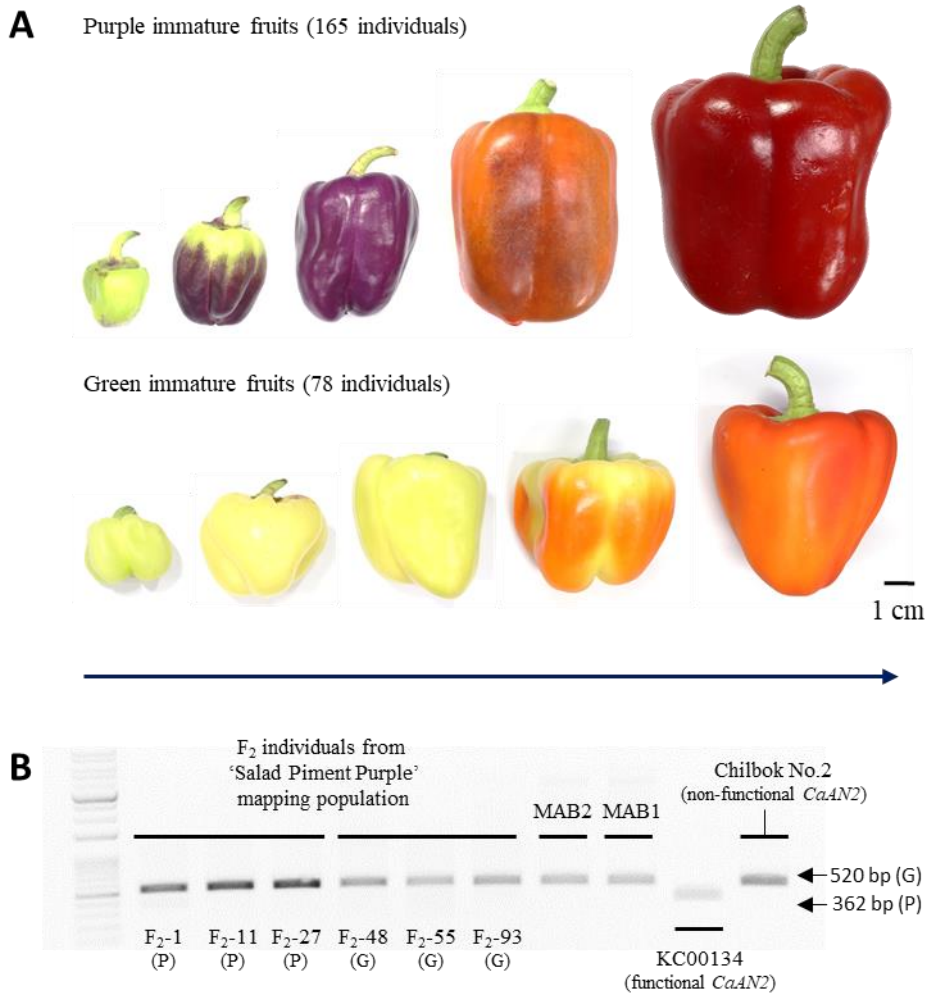

Fig. 2. Phenotype and genotype analysis of the mapping population. (A) Fruit phenotypes of the F_2 plants derived from *C. annuum* 'Salad Piment Purple'. Pigmentations in immature fruits were segregated into purple and green. (B) *CaAN2* genotype analysis of the mapping population. Plants in the mapping population had the non-functional *CaAN2* allele, but still segregated for purple pigmentation. Previously developed SCAR marker set was used for screening *CaAN2* functionalities. Two *C. annuum* lines were used as controls: 'KC00134' and 'Chilbok No.2' for *CaAN2* functional and non-functional allele, respectively.

The *CaAN3* locus was located on chromosome 10

To identify the gene *CaAN3*, RNA sequencing analysis was combined with BSA. Three RNA pools were prepared from six individual plants for each phenotypic bulk (18 purple and 18 green plants), and sequenced using the Illumina platform. An average of 121,311,533 reads was produced for each of the RNA pools by BSR-Seq, and total read bases per a pooled RNA were 33.6 Gbp for purple and 39.9 Gbp for green bulk, with the coverage of 748× and 890× for each bulk. When raw sequences were aligned to the reference genome Dempsey v1.0 (unpublished), an average of 340 million reads was mapped, covering about 96% of total coding region length (Table 2). A total of 311,679 raw SNPs were called, and 63,316 SNPs were obtained after filtering out the non-polymorphism and low quality value. Using these filtered SNPs, Δ(SNP – index) values, which mean the SNP differences in mapped reads between the two different pools, were calculated.

To confirm the candidate region and delimit the *CaAN3* locus, 6,672 SNPs were selected with Δ(SNP – index) values higher than the 99% confident level. These significant SNPs were analyzed, and five candidate regions were selected locating on chromosomes 1, 8, and 10. The highest peak of Δ(SNP – index) values was detected on chromosome 10. Two candidate regions were found on chromosome 10, one located from 0.15 to 108.8 Mb, and the other located from 121.7 to 239.8 Mb (Fig. 3).

Table 2. Summary of BSR-Seq analysis for purple and green bulks

Information	Purple	Green
No. of individuals in pools	18	18
No. of raw reads	332,378,020	395,491,780
Total size (Gbp)	33.6	39.9
Coverage	~ 748×	~ 890×
Mapped reads number	310,907,661	369,932,886
Mapped reads percentage (%)	96.1	96.1

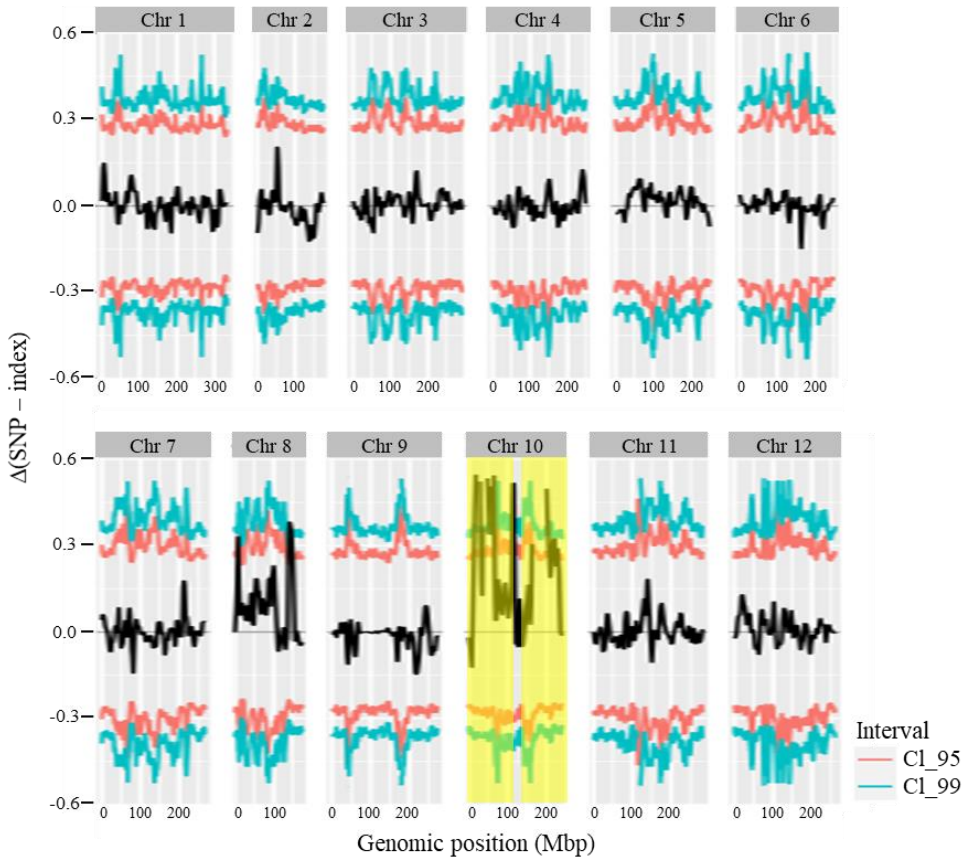

Fig. 3. Estimation of the candidate region for *CaAN3*. A total of 6,672 significant SNPs were identified by calculating Δ(SNP – index) between two BSR-Seq data. Yellow boxes indicate the putative genomic region where candidate gene is expected to be located at.

Several anthocyanin biosynthetic genes expressed differentially in purple and green phenotypes

Since we were unable to pinpoint the *CaAN3* location on chromosome 10 by Δ(SNP – index), differentially expressed genes (DEGs) was analyzed by comparing the transcriptome data of purple and green pools obtained from BSR-Seq. As a result, a total of 2,175 significant DEGs were found. Gene ontology (GO) term enrichment analysis using 2,175 DEGs showed that twelve GO terms were significantly enriched, including secondary metabolic process and carbohydrate metabolic process categorized as 'biological process', transcription factor activity and catalytic activity categorized as 'molecular function', and extracellular region categorized as 'cellular component' (Fig. 4A). In addition, eight Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were observed, and four of them were related to anthocyanin biosynthesis pathways such as 'biosynthesis of secondary metabolites', 'metabolic pathways', 'phenylpropanoid biosynthesis', and 'flavonoid biosynthesis' (Fig. 4B).

When analyzing the expression levels of the anthocyanin biosynthetic genes following the pathway starting from phenylalanine (Fig. 5A), *PAL*, *C4H*, *4CL*, *CHS*, *CHI*, and *F3H*, which are classified as EBGs, were not differentially expressed between purple and green pools. In contrast, there were significant differences of expression levels of *DFR*, *ANS*, and *3GT*, classified as LBGs, between purple and green pools (Fig. 5B).

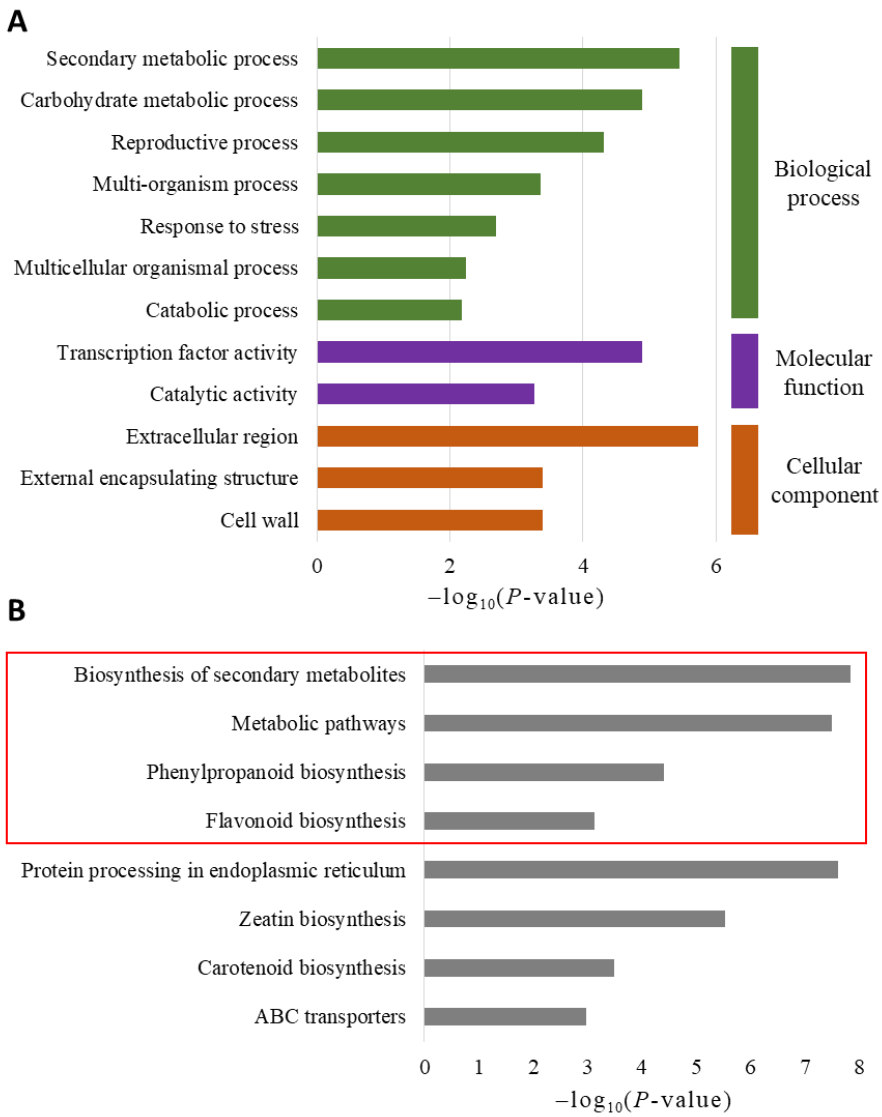

Fig. 4. Transcriptome analysis using fruit-specific DEGs. (A) GO term enrichment analysis; GO analysis was performed in three categories, biological process, molecular function, and cellular component. (B) KEGG pathway enrichment analysis. Pathway enrichment analysis using 2,175 DEGs involved in fruit-specific anthocyanin accumulation. The red box indicates known anthocyanin biosynthesis related pathways.

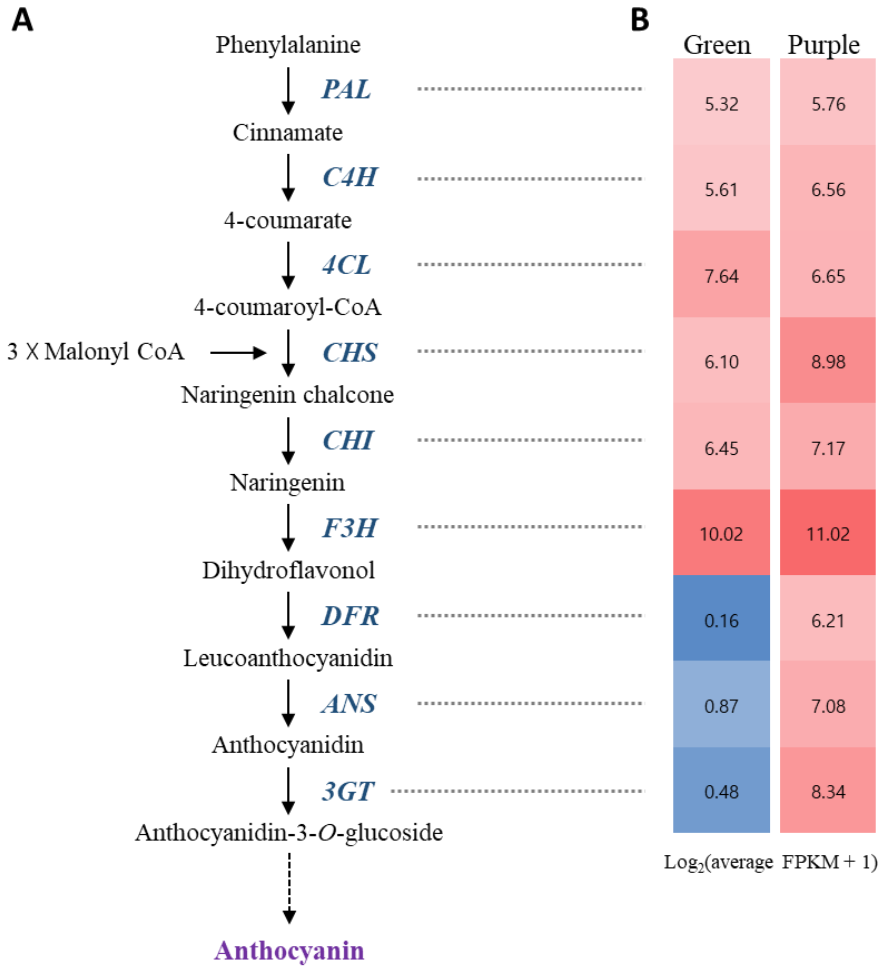

Fig. 5. Expression of anthocyanin biosynthetic genes in the pericarp of green and purple pools from 'Salad Piment Purple' F_2 population. (A) Core anthocyanin biosynthesis pathway in pepper. *PAL*, phenylalanine ammonialyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, 4-coumaroyl CoA; *CHS*, chalcone synthase; *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 anthocyanin biosynthetic pathway. The numbers in the boxes indicate the expression level of the genes, values of Log_2(average FPKM + 1), of each phenotypic pool.

Dem.v1.00043895 was the strongest candidate gene for *CaAN3*

In the BSR-Seq analysis, the physical position of the *CaAN3* locus was estimated to locate at 0.15 - 108.8 Mbp and 121.7 - 239.8 Mbp regions on chromosome 10. When 'Salad Piment Purple' F_2 population was subjected to genotyping with HRM-10-63 at 62.9 Mbp and SWPm_00416 at 168.1 Mbp, recombinants were observed demonstrating that the target region should be repositioned around the 168.1 Mb region on chromosome 10. A total of five markers, which have additionally been developed based on SNPs information obtained from BSR-Seq, were used to fine map the *CaAN3* region. Based on the recombination patterns of molecular markers, *CaAN3* could be located at the 184.6 - 186.4 Mbp region chromosome 10 of Dempsey v1.0 reference genome (Fig. 6).

In this region, fifteen candidate genes were annotated. The DEG analysis revealed that eight candidate genes were not expressed in both purple and green bulks. Three candidates, *Dem.v1.00043888*, *0043915*, and *0043916* showed minimal expression levels ranging from 0.01 to 0.41 Log_2(average FPKM + 1) values. Another three candidates, *Dem.v1.00043885*, *0043901*, and *0043918* showed relatively high expression levels, but there were no significant differences in expression levels between purple and green bulks. *Dem.v1.00043895*, which was annotated as a MYB transcription factor, showed no expression in the green bulk but significantly higher expression levels in the purple bulk (Table 3), demonstrating *Dem.v1.00043895* is the strongest candidate gene for *CaAN3*.

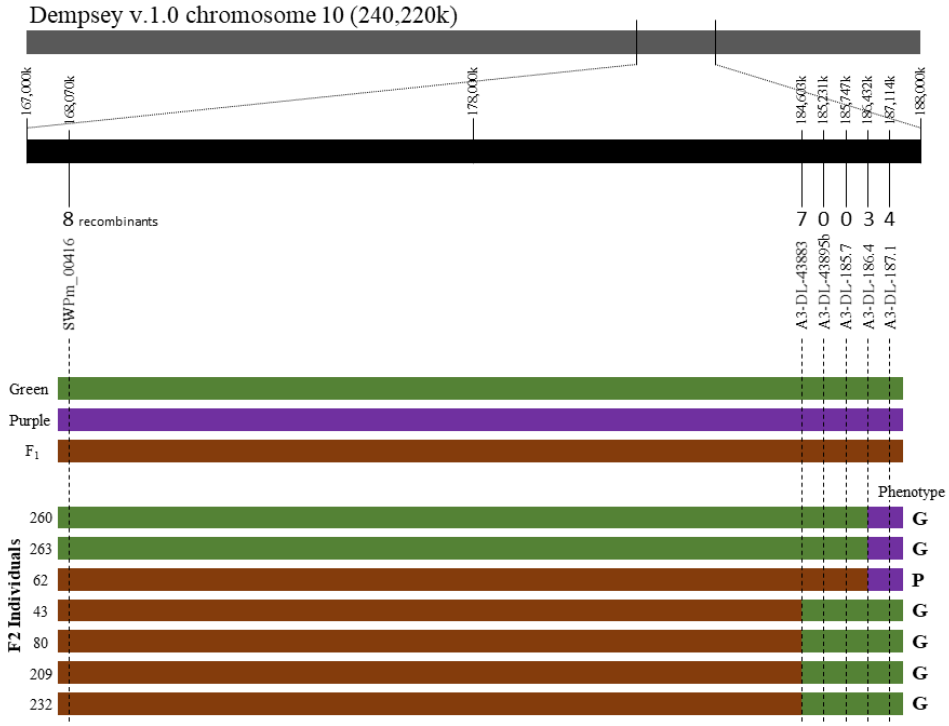

Fig. 6. Fine mapping of *CaAN3*. The *CaAN3* region was mapped on chromosome 10, ranging from 184,603kb to 186,432kb. The number of recombinants from 243 F_2 individuals are marked with number below the black bar. Genotypes of green homozygotes, purple homozygotes, and heterozygous 'Salad Piment Purple' cultivar are respectively indicated by green, purple, and brown bars. Seven bars at the bottom represent observed genotypes of individuals with recombination.

Table 3. Annotated genes in the *CaAN3* region on chromosome 10 and relative gene expression levels in DEGs analysis

Gene	Physical location (Mbp)	Annotation	Relative gene expression Log_2(average FPKM + 1)	
			Green	Purple
DEM.v1.00043883	184.6	Similar to MYB1: Transcription factor MYB1 (*Actinidia chinensis* var. *chinensis*)	0.00	0.00
DEM.v1.00043884	184.7	Protein of unknown function	0.00	0.00
DEM.v1.00043885	184.8	Heavy metal-associated isoprenylated plant protein 6 (*Arabidopsis thaliana*)	6.72	6.69
DEM.v1.00043888	185.0	Protein of unknown function	0.00	0.41
DEM.v1.00043895	185.2	Similar to MYB1: Transcription factor MYB1 (*Actinidia chinensis* var. *chinensis*)	0.00	6.10
DEM.v1.00043899	185.6	Similar to MYB113: Transcription factor MYB113 (*Arabidopsis thaliana*)	0.00	0.00
DEM.v1.00043901	185.7	Similar to ABC1K1: Protein ACTIVITY of BC1 complex kinase (*Arabidopsis thaliana*)	4.33	4.23
DEM.v1.00043902	185.9	Similar to ALS3: Protein ALUMINUM SENSITIVE 3 (*Arabidopsis thaliana*)	0.00	0.00
DEM.v1.00043903	186.0	Similar to HD1: Homeobox protein HD1 (*Brassica napus*)	0.00	0.00
DEM.v1.00043908	186.1	Protein of unknown function	0.00	0.00
DEM.v1.00043909	186.1	Protein of unknown function	0.00	0.00
DEM.v1.00043910	186.1	Similar to MYB1: Transcription factor MYB1 (*Actinidia chinensis* var. *chinensis*)	0.00	0.00
DEM.v1.00043915	186.2	Protein of unknown function	0.00	0.06
DEM.v1.00043916	186.2	Similar to NAC078: NAC domain-containing protein 78 (*Arabidopsis thaliana*)	0.01	0.03
DEM.v1.00043918	186.4	Similar to UGT86A1: UDP-glycosyltransferase 86A1 (*Arabidopsis thaliana*)	9.53	10.4

The promoter region of *Dem.v1.00043895* had structural variation

To reveal sequence variations of *Dem.v1.00043895* between purple and green plants, individuals randomly selected from 'Salad Piment Purple' F_2 population were sequenced for the promoter and coding regions. Leaves from individuals with homozygous purple and green genotypes as well as 'MAB2' (purple immature fruit) and 'MAB1' (green immature fruit) were sampled and sequenced. The sequencing results of the promoter region showed that F_2 individuals with purple homozygotes and 'MAB2' had identical promoter structures. Likewise, F_2 individuals with green homozygotes and 'MAB1' had identical promoter structures.

In the promoter region, several structural variations were found between the purple and green *CaAN3* alleles. There were 4 bp insertion at the position -1,428, 7 bp insertion at -1,069, 35 bp insertion at -887, and 15 bp deletion at the position -157 in the promoter region of the green *CaAN3* allele (Fig. 7A). Meanwhile, protein sequence of coding region suggested that there was an A to D variation between purple and green at 43rd amino acid. A variation was also found at 211th amino acid, but there was no consistency between F_2 purple homozygotes and 'MAB2' (Fig. 7B). Considering these sequence variations in relation to the gene expression pattern, it could be hypothesized that the sequence variation in the promoter region might cause no expression of *Dem.v1.00043895*. In other words, while the purple *CaAN3* allele is functional, the green allele is non-functional due to the structural difference of the promoter region.

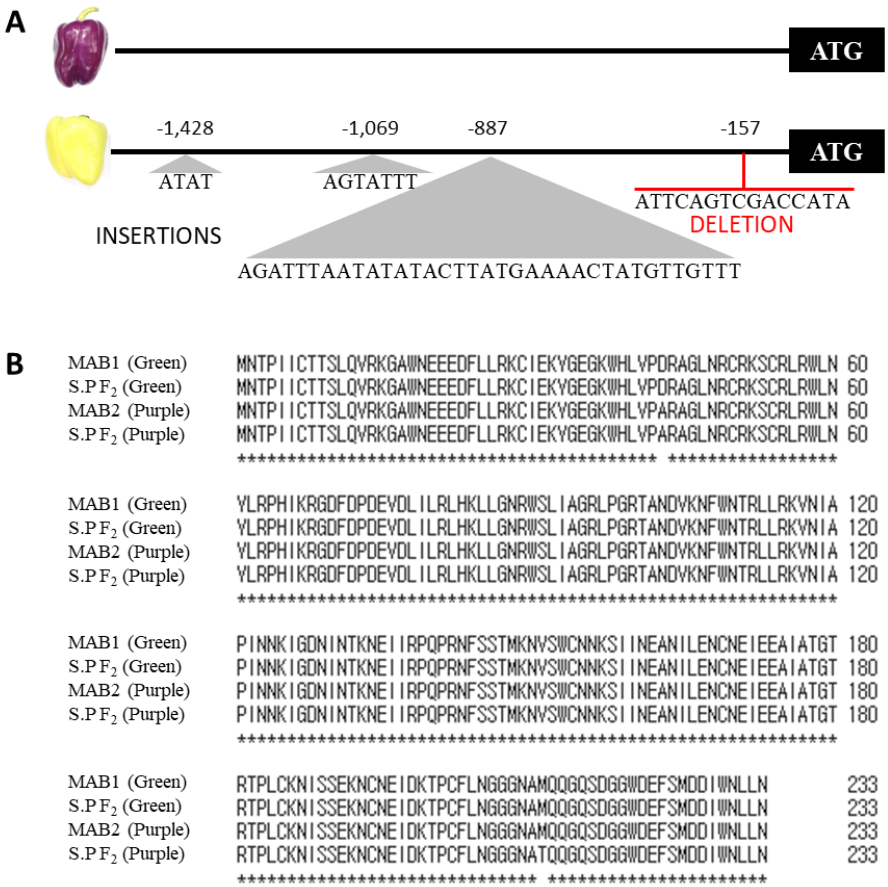

Fig. 7. Structural variation between purple and green *CaAN3* alleles. (A) Major sequence variation between promoters of putative functional and non-functional alleles. Purple homozygous individuals of 'Salad Piment Purple' F_2 and 'MAB2' have identical sequences. Green homozygous individuals of 'Salad Piment Purple' F_2 and 'MAB1' have identical sequences. Due to the expression pattern of *Dem.v1.00043895*, Purple allele which expressed normally was hypothesized as 'functional' and Green allele was hypothesized as 'non-functional' (B) Comparison of protein sequence between two purple alleles and two green alleles. Premature termination of protein translation did not happen.

Dem.v1.00043895 was specifically expressed only in the immature fruits

To investigate the expression pattern of the candidate gene *Dem.v1.00043895*, 'MAB2' fruits were sampled according to developmental stages (Fig. 8A) and the expression analysis was performed. The expression analysis revealed that the gene showed the highest relative expression level at the beginning of pigmentation in the very early stages of fruit. As the fruit matured and the purple pigmentation disappeared, relative expression levels gradually decreased. After 40 days post anthesis (DPA) when fruits lost most of their purple pigment and turned into mature red color, the gene showed almost no expression (Fig. 8B). Anthocyanidin concentration measured by HPLC analysis showed the similar pattern with the pattern of phenotypic changes of the pericarp. The highest anthocyanidin per gram of dry weight was measured in the earliest stage of growth, and it showed a tendency to gradually decrease as fruit develops. Five different anthocyanidin (cyaniding, delphinidin, malvidin, pelargonidin, and peonidin) were measured by HPLC analysis, and more than 99% of the total anthocyanidin contents were found to be delphinidin (Fig. 8C). When the expression of *Dem.v1.00043895* was investigated by tissues, this gene was confirmed to be specifically expressed in the immature fruits at 20 DPA, whereas the gene was barely expressed in leaves, stems, and flowers (Fig. 8D).

Silencing of *Dem.v1.00043895* inhibited purple pigmentation in pericarp

To test whether *Dem.v1.00043895* functions as *CaAN3*, the TRV-mediated VIGS approach was applied. The coding region from *Dem.v1.00043895* was cloned

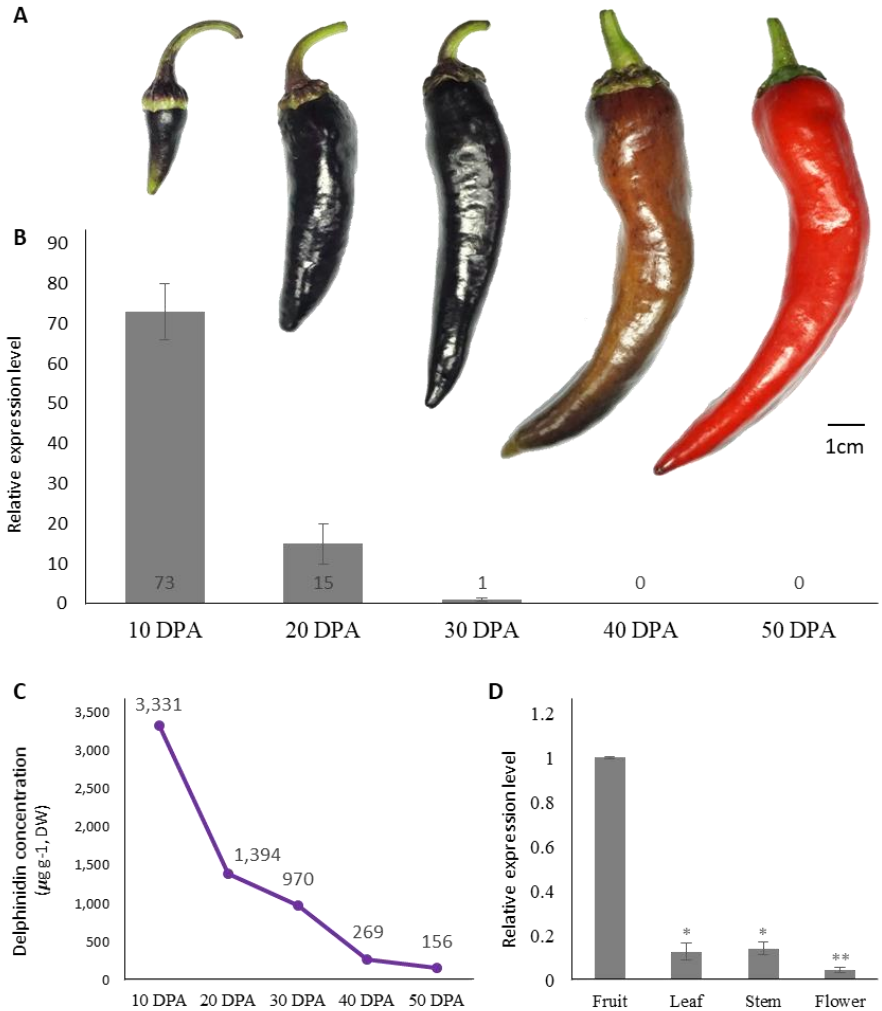

Fig. 8. Physiological changes of 'MAB2' fruits during development. (A) Phenotypic changes of fruit pigmentation from 10 to 50 DPA. (B) Relative expression levels of *Dem.v1.00043895* according to the fruit development. (C) Changes in concentration of delphinidins in the pericarp. Most of the anthocyanidins measured were delphinidins. (D) Relative expression levels of *Dem.v1.00043895* in different tissues. *Dem.v1.00043895* showed significantly lower expression levels in leaves, stems, and flowers than in immature fruit at 20 DPA.

into gene silencing TRV2 vector. *Agrobacterium* carrying pTRV2::GFP and pTRV2::PDS was used as negative and positive controls, respectively. After *Agrobacterium* infiltration to *C. annuum* 'MAB2', neither showed alteration in fruit pigmentation, whereas pTRV2::PDS inoculated plants showed bleaching leaves indicating that PDS was properly silenced. Both pTRV2::GFP and pTRV::PDS inoculated plants showed fully purple fruits which were same as wildtype 'MAB2'. However, in plants inoculated by TRV2 vector harboring *Dem.v1.00043895*, alterations of fruit pigments were observed. Although inoculated plants showed slight differences in the phenotypes, most of the inoculated plants showed silenced phenotype, no accumulation of anthocyanin. In some plants, immature fruits stayed in green instead of purple with most of their pericarps in their immature stages. Other plants showed fruits with a partially purple and green pericarp. One of the characteristics of these fruits was that the borders between purple and green parts were quite clearly distinguished (Fig. 9A). It was not observed that the green part of the pericarp never turned into purple again as the stages of development continued. The HPLC analysis was performed using several fruits from several different plants, including fruits in both immature and mature stages. The fruits showing similar silencing phenotypes were randomly selected and pooled. The result confirmed that delphinidins were not actually accumulated in the regions where the purple pigmentation was not occurred (Fig. 9B).

The gene expression analysis for anthocyanin biosynthesis genes revealed that when *Dem.v1.00043895* was silenced, significant downregulation of expression of

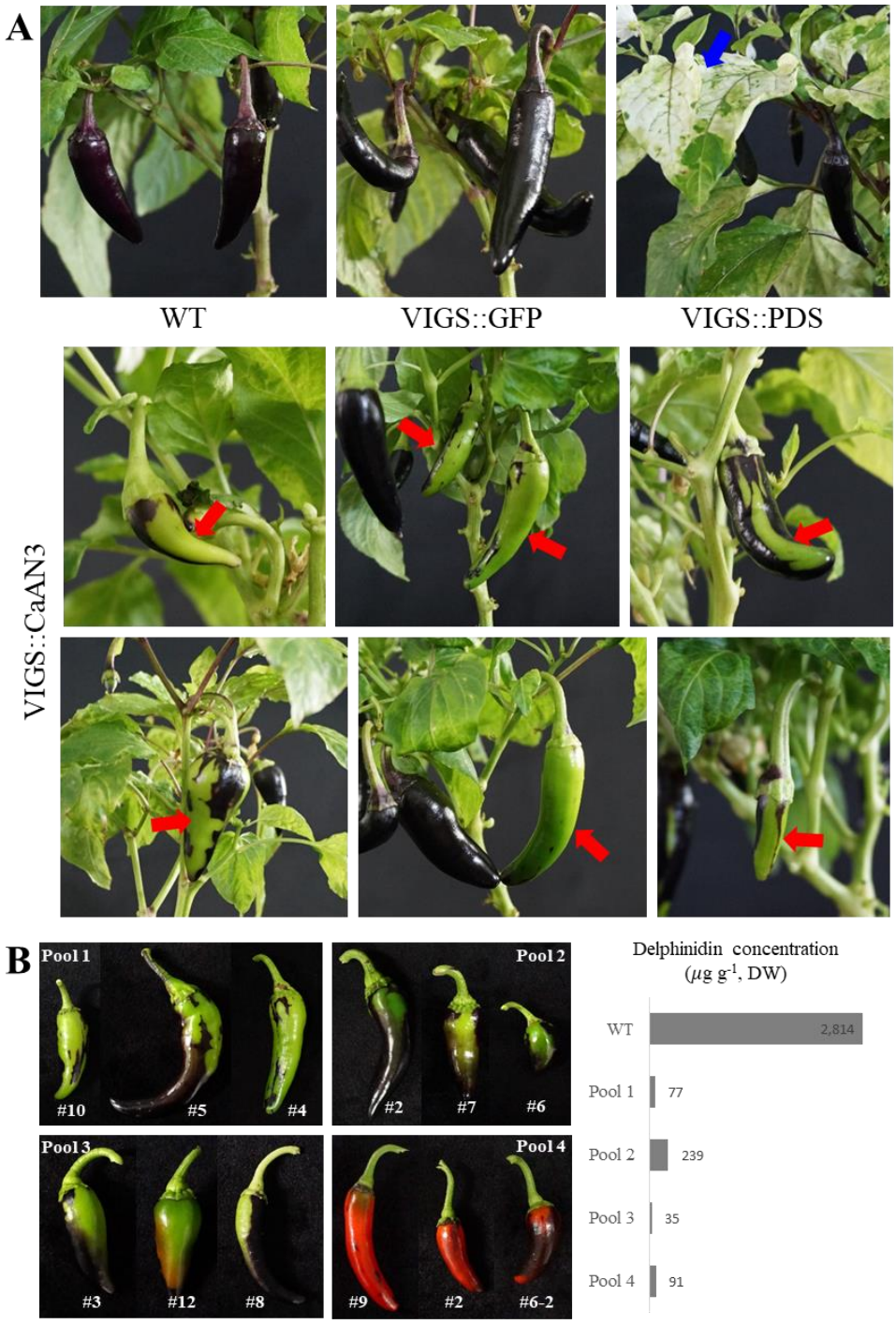
A
WT
VIGS::GFP
VIGS::PDS
VIGS::CaAN3
B
Pool 1
#10
#5
#4
Pool 2
#2
#7
#6
Pool 3
#3
#12
#8
Pool 4
#9
#2
#6-2
Delphinidin concentration
(μg g-1, DW)
WT
2,814
Pool 1
77
Pool 2
239
Pool 3
35
Pool 4
91

Fig. 9. Alteration of fruit pigment by silencing *Dem.v1.0.00043895*. VIGS study showed the knock-out of *CaAN3* candidate gene causes diminishment or elimination of purple pigment. (A) Phenotypes of fruits of WT, VIGS::GFP, VIGS::PDS, and VIGS::CaAN3 plants. VIGS::GFP plants, and VIGS::PDS plants were used as negative and positive controls, respectively. VIGS::GFP and VIGS::PDS plants showed only purple fruits as wildtype. But when *CaAN3* was silenced, purple pigment was partly diminished or eliminated to some degree. Blue arrow indicates bleached leaf suggesting PDS gene was properly silenced. Red arrows indicate loss of purple pigment in *CaAN3* silenced fruits. (B) HPLC analysis of anthocyanidin concentration of 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 the mature fruits (pool 4). Most of the anthocyanidins measured were delphinidins.

several pathway genes occurred. Fruits from pTRV::GFP inoculated plants were used as negative control. Both EBGs and LBGs showed significant downregulation of expression. However, there was no significant difference in the expression levels of *4CL* and *F3H* in EBGs between the gene-silenced and control fruits (Fig. 10A). LBGs showed a more obvious differences. *DFR*, *F3'5'H*, and *ANS* all showed significant downregulation in expression, when *Dem.v1.00043895* was silenced. In particular, expression levels of *DFR* and *F3'5'H* were barely detectable in the *Dem.v1.00043895* silenced fruits (Fig. 10B). Based on this result, Dem.v1.00043895 could be hypothesized to directly regulate LBGs in anthocyanin biosynthetic pathway.

Overexpression of *Dem.v1.00043895* caused accumulation of anthocyanin in *N. benthamiana* leaves

To further test whether *Dem.v1.00043895* can function as *CaAN3*, transient overexpression assay in *N. benthamiana* leaves was performed. The coding region from *Dem.v1.00043895* was cloned into the pCAMBIA2300 vector under 35S promoter (Fig. 11A), and infiltrated into young *N. benthamiana* leaves. *Agrobacterium* carrying pCAMBIA2300::GFP was served as a negative control. About 10 days after the inoculation, the alteration of color was visually observed around the inoculation site, while leaves inoculated with pCAMBIA2300::GFP showed no alteration in color (Fig. 11B). When the leaves were sampled and anthocyanin accumulation was tested (Mazzucato et al., 2013), leaves inoculated by

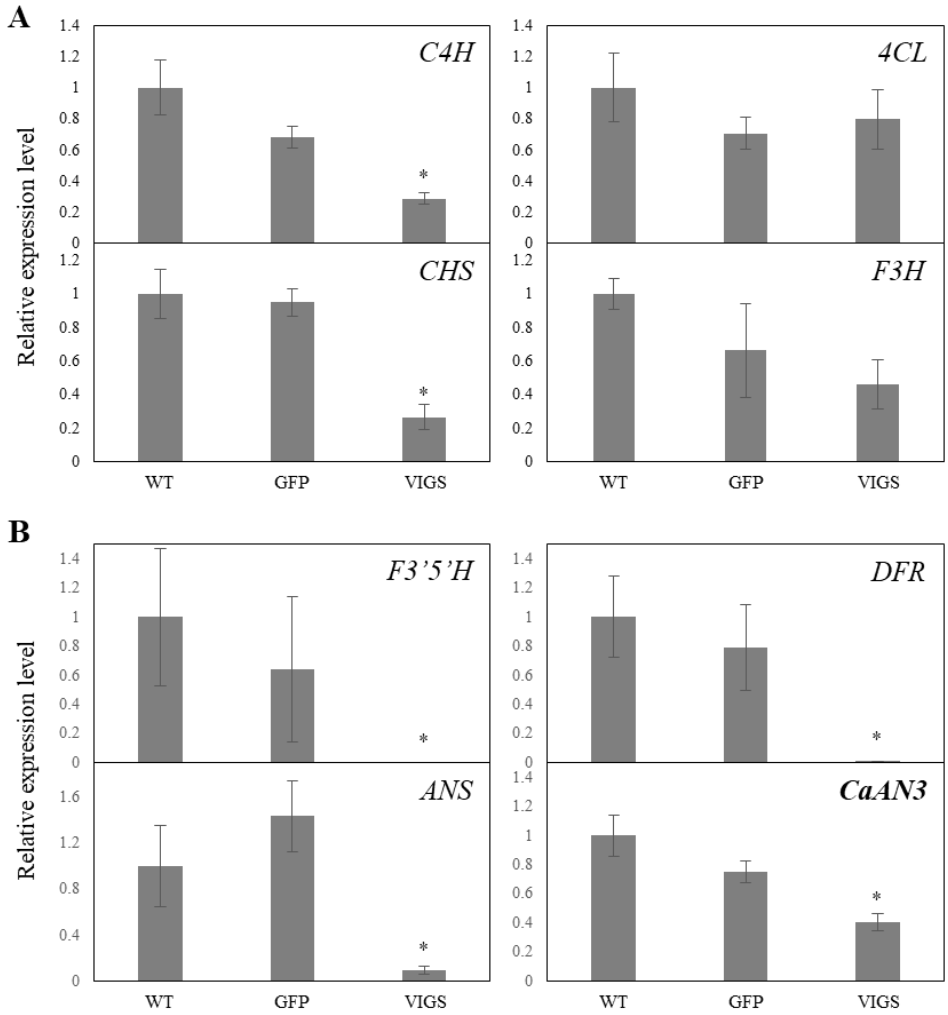

Fig. 10. Relative gene expression in the pericarp of *Dem.v1.00043895* silenced fruits. (A) Relative expression levels of four EBGs (*C4H*, *4CL*, *CHS*, *F3H*) in the pericarps of wildtype, *GFP* silenced fruits as negative control, and *Dem.v1.00043895* silenced fruits. (B) Relative expression levels of three LBGs (*F3'5'H*, *DFR*, *ANS*) and *CaAN3* candidate gene. Star markers indicate significant significance in difference at the 95% confident level.

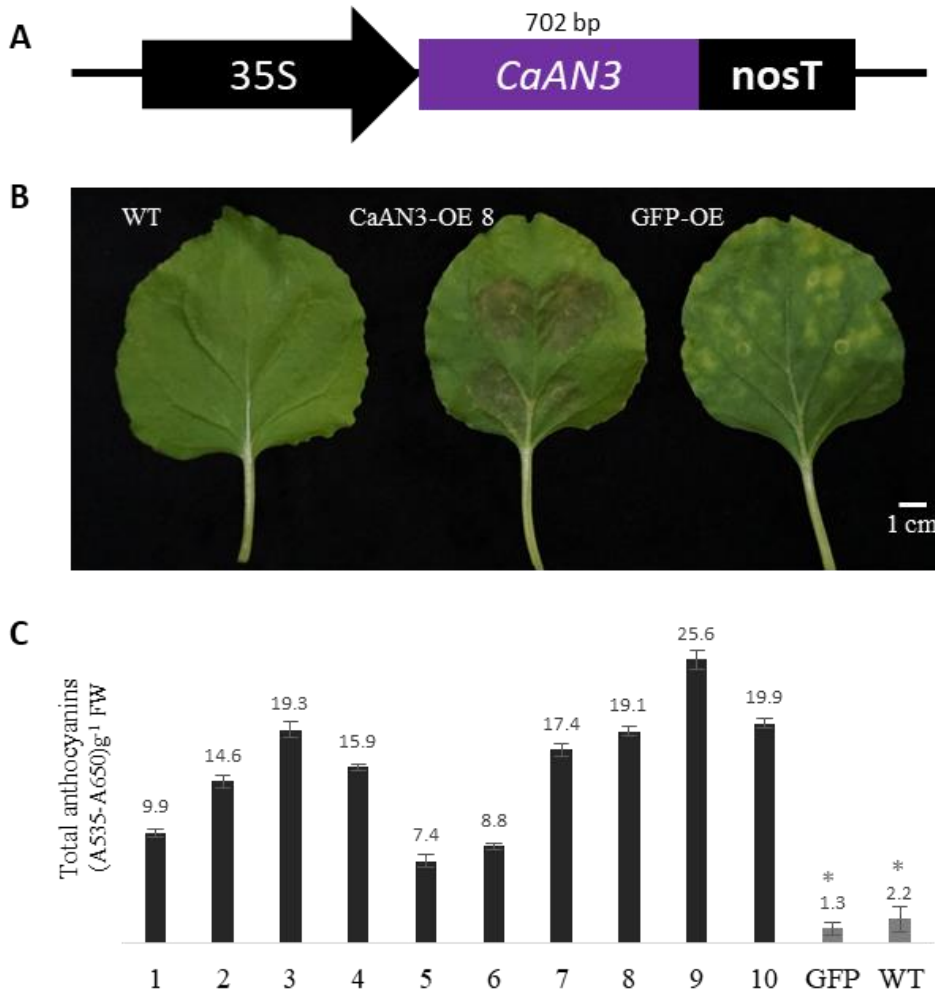

Fig. 11. Transient overexpression of *CaAN3* in *N. benthamiana* leaves. (A) *CaAN3* construct containing CaMV 35S overexpression promoter. (B) Phenotypic alteration of *CaAN3* overexpressed *N. benthamiana* leaf, and *GFP* overexpressed leaf for negative control. (C) Total anthocyanin contents measured by spectrophotometer. Each leaf inoculated by overexpression vector showed the anthocyanin accumulation to some extent, while *GFP* overexpressed leaves and wildtype leaves showed no accumulations.

Dem.v1.00043895 overexpression vector showed clear purple color, while wildtype and negative control showed only slightly green color. When the quantification analysis of anthocyanin via absorbance measurement using a spectrophotometer was performed, *Dem.v1.00043895* overexpressed leaves showed the (A535-A650)/g levels varying from 7.4 to 25.6 according to the intensity of the overexpression, which showed the same tendency similar to the depth of purple that could be discerned visually. WT or *GFP* overexpressed leaves showed significantly lower anthocyanin levels than *Dem.v1.00043895* overexpressed leaves (Fig. 11C).

The expression patterns of *CaAN2* and *CaAN3* had a correlation with phenotypes of fruits and flowers in pepper

To test the hypothesis that *Dem.v1.00043895* is a master regulator of fruit-specific anthocyanin biosynthesis, and that the structural variation of the promoter region determines the functionality of this gene, several pepper accessions with purple immature fruit and some control accessions with green immature fruits were evaluated. ‘MAB2’ and ‘MAB1’ were also included in the group to set the standard. Among 24 different pepper accessions, four different phenotypic patterns were observed and categorized into four groups: Group I including eight accessions having small purple fruits and purple flower, Group II consisted of four accessions including ‘MAB2’ having purple fruits and white flower (fruit-specific pigmentation), Group III containing four accessions showing pale purple pigmentation while immature fruits having yellowish background coloration, and Group IV consisted of

eight accessions including ‘MAB1’ having green immature fruit with white flowers (Fig. 12). More detailed phenotypes of 24 pepper accessions including leaf and stem are summarized in Table 4.

Next, expression levels of the candidate gene *CaAN3* and known *CaAN2* were analyzed for these 24 different accessions. Several consistent trends were observed. First, the expression levels of *CaAN2* showed a positive correlation with accessions in Group I with fruit and flowers with purple coloration. Second, *CaAN3* was only expressed in accessions in Group II with fruit-specific purple pigmentation, as hypothesized. Four accessions in Group III with yellow immature skin color with partial pale purple pigmentation showed little *CaAN2* expression and almost none of *CaAN3* expression. And accessions in Group IV with green immature fruit showed almost no expression of both *CaAN2* and *CaAN3* (Fig. 13). Notably, *CaAN2* was expressed only in pepper accessions with purple flowers. Three of these accessions also showed purple pigmentation in leaves.

Structural variations in the promoter region were not directly related to *CaAN3* expression

A SCAR marker set was developed targeting the indel mutation found at the promoter region (Table 5). To genotype the 24 accessions, the SCAR marker was used to amplify promoter regions of the lines. When the SCAR marker analyzed genotype of 24 accessions, the ‘MAB1’ accession, which has the same structural variation just like the green *CaAN3* allele of ‘Salad Piment Purple’, showed the

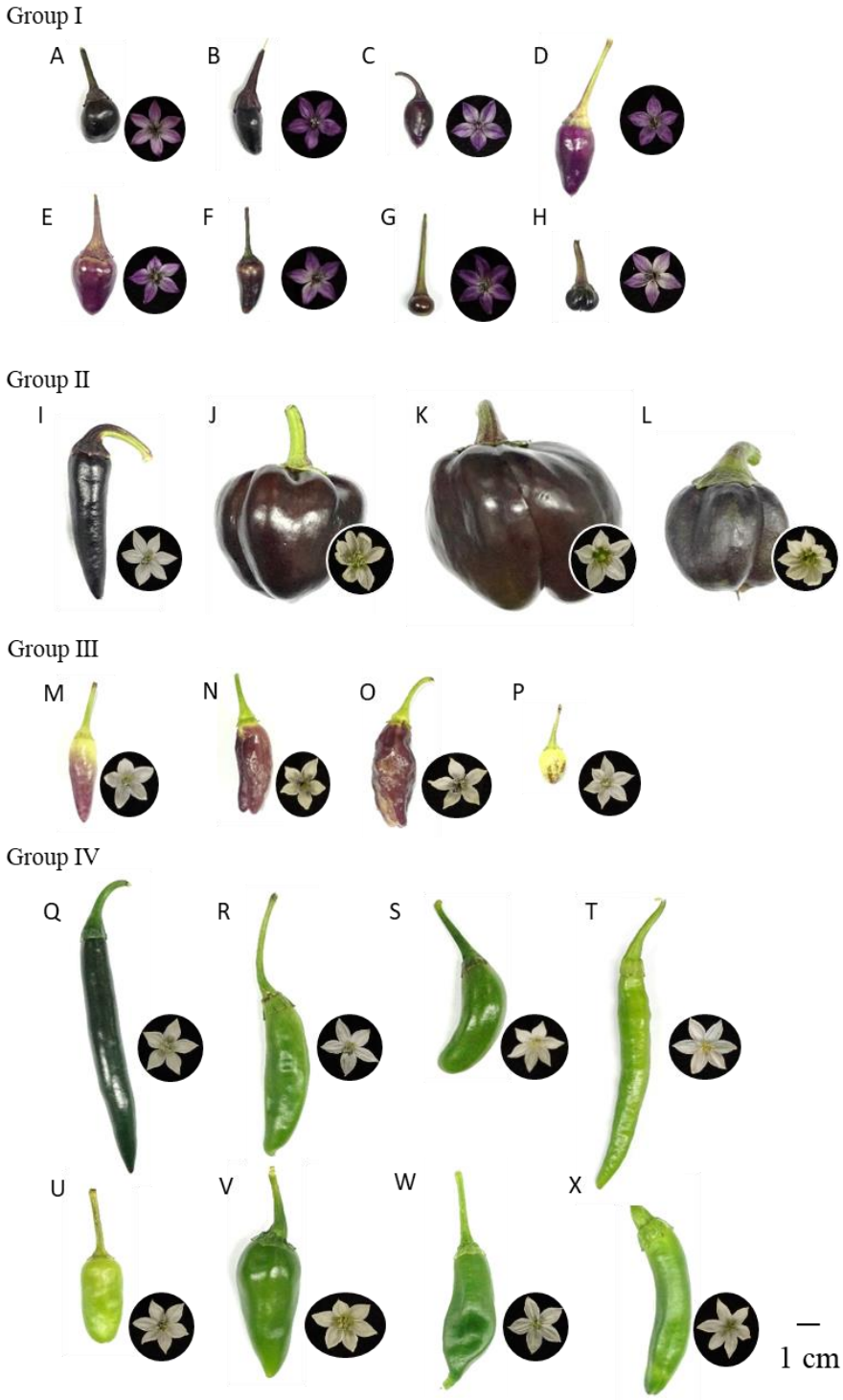
Group I
A
B
C
D
E
F
G
H
Group II
I
J
K
L
Group III
M
N
O
P
Group IV
Q
R
S
T
U
V
W
X
1 cm

Fig. 12. Phenotypes of immature fruits and flowers of the 24 pepper accessions. Four different phenotypic patterns were observed. (A-H) showed both purple fruits and purple pigmented flowers. (I-L) showed fruit-specific purple pigmentation. (M-P) showed pale purple pigmentation with yellowish background. (Q-X) showed no purple pigmentation on both fruit and flower. (A) IT218962. (B) KC00134. (C) AC08-181. (D) IT223782. (E) IT286136. (F) IT218615. (G) IT301140. (H) IT158844. (I) MAB2. (J) Mavis. (K) Mavras. (L) IT305471. (M) IT286162. (N) IT158637. (O) IT229203. (P) AC09-003. (Q) KC01237. (R) IT218898. (S) Chilbok No.2. (T) Jeju. (U) IT238052. (V) Yuwolcho. (W) IT238056. (X) MAB1.

Table 4. Phenotypes of the 24 pepper accessions.

Group	No. (Fig. 12)	Species	Accession number	Phenotype*				
				Immature Fruit	Mature Fruit	Flower	Leaf	Stem
Ⅰ	A	*C. frutescens*	IT218962	P	R	P	P	G
	B	*C. annuum*	KC00134	P	R	P	P	G
	C	*C. annuum*	AC08-181	P	R	P	P	G
	D	*C. annuum*	IT223782	P	R	P	G	G
	E	*C. annuum*	IT286136	P	R	P	G	G
	F	*C. annuum*	IT218615	P	R	P	G	G
	G	*C. frutescens*	IT301140	P	R	P	G	G
	H	*C. annuum*	IT158844	P	R	P	G	G
Ⅱ	I	*C. annuum*	MAB2	P	R	W	G	G
	J	*C. annuum*	Mavis	P	R	W	G	G
	K	*C. annuum*	Mavras	P	R	W	G	G
	L	*C. annuum*	IT305471	P	R	W	G	G
Ⅲ	M	*C. annuum*	IT286162	P’	R	W	G	G
	N	*C. chinense*	IT158637	P’	R	W	G	P
	O	*C. chinense*	IT229203	P’	R	W	G	G
	P	*C. annuum*	AC09-003	P’	R	W	G	G
Ⅳ	Q	*C. annuum*	KC01237	G	R	W	G	P
	R	*C. annuum*	IT218898	G	R	W	G	G
	S	*C. annuum*	Chilbok No.2	G	R	W	G	G
	T	*C. annuum*	Jeju	G	R	W	G	G
	U	*C. annuum*	IT238052	G	R	W	G	G
	V	*C. annuum*	Yuwolcho	G	R	W	G	G
	W	*C. annuum*	IT238056	G	R	W	G	G
	X	*C. annuum*	MAB1	G	R	W	G	G

* P, Purple; P’, Pale purple; G, Green; R, Red; W, White

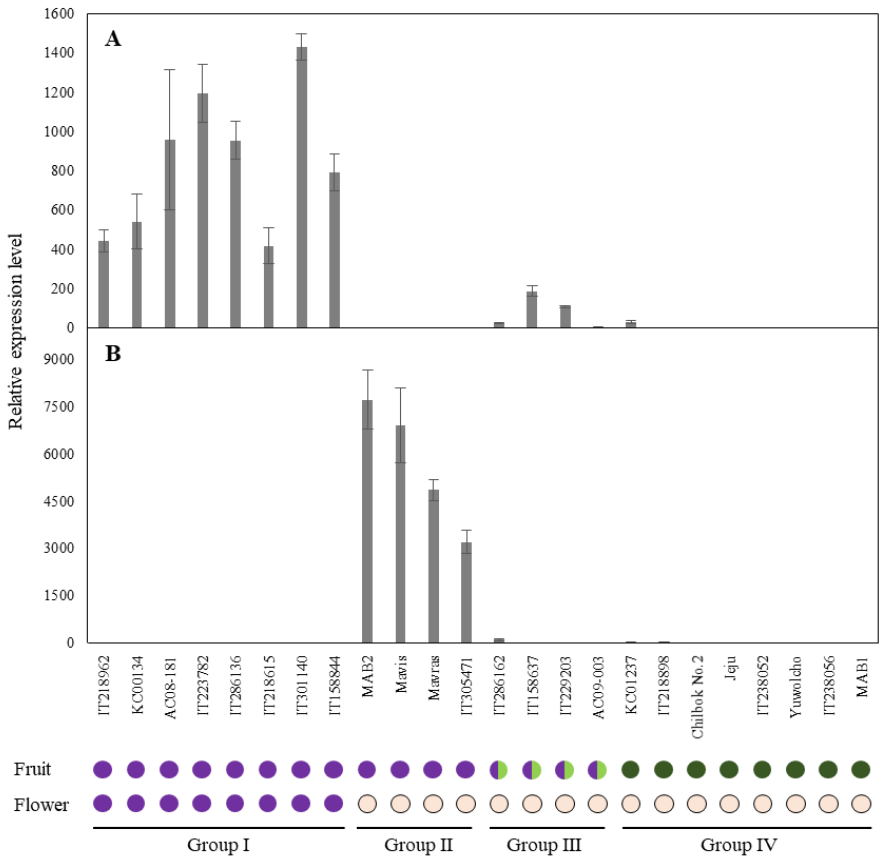

Fig. 13. Expression analysis of *CaAN2* (A) and *CaAN3* (B) in different pepper accessions. There were differences in the gene expression levels of each lines and a clear tendency in expression could be identified. Group I (from IT218962 to IT158844) with both purple fruits and flowers showed relatively high level of *CaAN2* expression but showed almost none of *CaAN3* expression. In case of Group II (from MAB2 to IT305471), *CaAN2* was barely expressed while *CaAN3* showed quite higher expression level than other lines. Group III with light yellow tones of immature fruits with pale purple pigmentations showed low *CaAN2* expression level and almost no *CaAN3* expression level (from IT286162 to AC09-003). And Group IV with no purple pigmentation on both fruits and flowers showed lowest expression level of *CaAN2* and *CaAN3*.

Table 5. A primer set for the SCAR marker targeting *CaAN3*

Primer	**Sequence (5'-3')**
AN3_IDEL_F	GCGGATCAGATTCATCTTTCATCTAG
AN3_INS_R	ATTGAAAACAACATACTTTTCATAAGTATATATTAAATCT
AN3_DEL_R	GGTATGGTCGACTGAATATCCTTCA

expected 524 bp of specific band. However most of the other lines showed genotype same as ‘MAB2’ regardless of variation in fruit and flower pigmentations, showing 1,188 bp of specific band (Fig. 14). Sequence analysis of two pepper accessions ‘IT218962’ and ‘KC00134’ with purple pigmentation on both fruit and flower, showed there is no structural differences in the promoter region of *CaAN3* comparing to ‘MAB2’ (data not shown) although it was proved to express only *CaAN2*. ‘IT158637’ and ‘IT229203’ accessions which have pale purple pigmentation on fruits showed relatively weak or no amplification in several attempts.

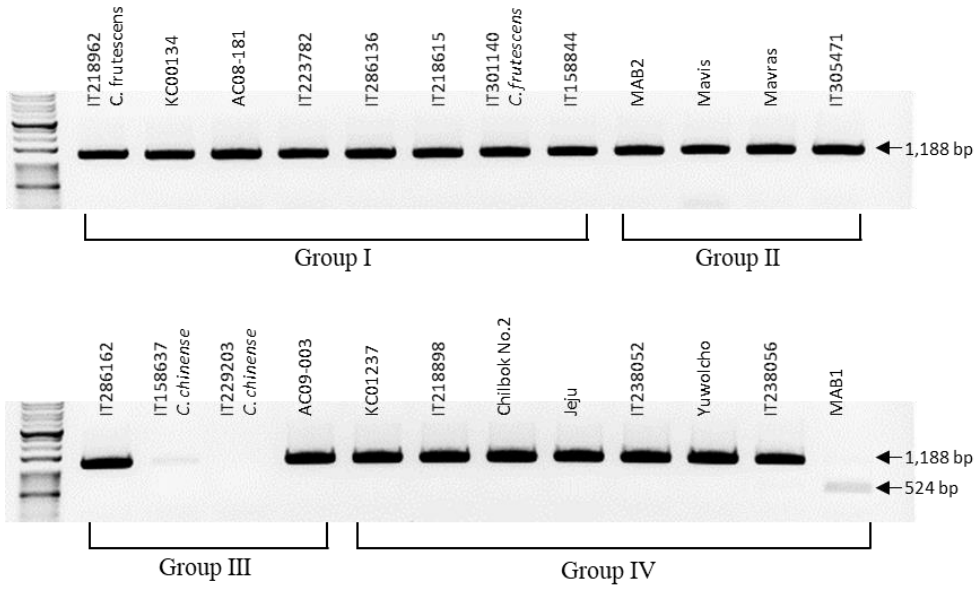

Fig. 14. Screening genotypes of the 24 accessions by using *CaAN3* SCAR marker. Although there are different expression patterns of *CaAN3* among 24 pepper lines, genotypic variation was not correlated. Except for 'MAB1', most of the other pepper lines have the identical structure as 'MAB2' with fruit-specific anthocyanin accumulation. 'IT229203' might has another type of promoter structure, since it had showed no specific amplifications in several attempts.

DISCUSSION

In this study, I identified that an R2R3 MYB transcription factor is encoded by the *CaAN3* gene and regulates the fruit-specific anthocyanin accumulation in pepper (*C. annuum)*. The gene annotated as *Dem.v1.00043895* in the reference genome of 'Dempsey v1.0' (unpublished) was determined to be *CaAN3*, which is located at 185.2 Mbp position on chromosome 10. Purple and green homozygotes in the F_2 population had distinct structural variations in their promoter regions. Furthermore, I confirmed the functional characteristics of *CaAN3*. Silencing *CaAN3* made it impair to synthesize fruit-specific anthocyanins in pepper, whereas overexpression of *CaAN3* enabled anthocyanin synthesis in *N. benthamiana*.

R2R3 MYB transcription factor is a well-known regulatory gene that functions as activator of structural genes in the anthocyanin biosynthesis pathway in a variety of plants (Jaakola, 2013). Several R2R3 MYB TFs acting as anthocyanin activators have also been reported in Solanaceous plants such as tomatoes, peppers, eggplants, and potatoes (Liu et al., 2018). However, relatively few studies have been conducted in pepper. In pepper, *CaAN2*, an R2R3 MYB TF, has been shown to positively regulate anthocyanin accumulation in tissues including immature fruits, flowers and leaves (Borovsky et al., 2004; Jung et al., 2019). The underlying mechanism was that the insertion of a non-LTR retrotransposon into the promoter determines the functionality of *CaAN2*. Comparison of the expression of structural genes involved

in anthocyanin biosynthesis between pepper accessions with functional and non-functional *CaAN2* alleles that the expression levels in *CaF3H*, *CaDFR*, *CaANS*, and *CaUFGT* were significantly downregulated in plants with non-functional alleles, suggesting *CaAN2* regulates several late biosynthesis genes of anthocyanin biosynthesis pathway (Jung et al., 2019). However, Jung (2019) showed that that there is another genetic factor regulating the expression of anthocyanin biosynthesis genes in addition to *CaAN2*, especially in accessions showing fruit-specific anthocyanin accumulation, because some pepper accessions showed the purple fruit phenotype despite having a non-functional *CaAN2* alleles.

Assuming that the novel locus *CaAN3* that regulates fruit-specific anthocyanin biosynthesis exists, BSR-Seq data was used to reveal the structural variations and transcriptomic differences of the purple and green plants at the breaker stage of fruits. At first, molecular markers had been developed based on SNPs information obtained from BSR-Seq to delimit the candidate region of *CaAN3*. When the expression analysis of structural genes in the anthocyanin biosynthesis pathway of fruit-specific purple and green fruit was performed, there were large differences in expression levels of late biosynthesis genes. One candidate gene was identified by investigating DEGs of all candidate genes located in the target region. Sequence analysis of the candidate gene and its promoter suggested there are several indel mutations in the promoter region of the green type allele and it might cause the loss of the functionality of *CaAN3*.

VIGS study and transient overexpression assay in *N. benthamiana* leaves

confirmed that *Dem.v1.00043895* actually has the function of regulating the expression of anthocyanin biosynthesis genes. Fruits of the gene-silenced plants lost purple pigments on immature stage and showed green immature fruit as other green fruited pepper accessions. Anthocyanin concentration of the pericarps was reduced in the same vein. Also, the gene combined with 35S overexpression promoter caused purple pigmentation in *N. benthamiana* leaves, confirming that anthocyanins were actually accumulated by overexpressing *CaAN3*.

However, when the genotype was conducted using several accessions through the SCAR marker set based on indel mutation on the promoter region, it was shown that most accessions, including accessions with green immature fruits, had the same structural variation as 'MAB2' having the functional allele, which is contradictory to the expected hypothesis from the research hypothesis that indel mutations in promoter region may be related to the expression of *CaAN3*. In particular, sequencing analysis of promoter regions of 'KC00134' and 'IT218962' accessions, which are expressing only *CaAN2* and not *CaAN3*, showed that the sequence was identical to that of 'MAB2' (data not shown). Therefore, it could be presumed that there is no direct correlation between the fruit-specific purple phenotype and the structural variation in the promoter region of the *CaAN3*.

Although this study was successful in identification of the *CaAN3* gene, an activator of fruit-specific anthocyanin accumulation, further studies are needed to reveal its expression mechanism. To study the mechanism underlying the phenomenon in depth, it is necessary to investigate the transcriptional network, that

is, how various transcription factors interact with each other in plants. In tomato, *SlAN2* and *SlANT1* encoding R2R3 MYB TFs that induce anthocyanin accumulation in fruit peel have been reported. It is well established that these two highly homologous genes co-express and co-segregate in plants and show significant correlation with anthocyanin biosynthesis in fruit peel (Povero et al., 2011). *SlAN2-like*, which encodes another R2R3 MYB activator, was recently proposed to operate as a master regulator of anthocyanin biosynthesis in tomato. This gene interacts with the bHLH TF encoded by *SlAN1*, to function as an MBW complex. However, it was shown that if the *SlMYBATV* gene encoding the R3 MYB repressor functions normally, it can competitively bind to *SlAN1* and inhibit *SlAN2-like* activity, thereby affecting the strength of anthocyanin accumulation (Sun et al., 2019).

In potato, three different R2R3 MYB activators of anthocyanin biosynthesis have been identified, namely *StAN1*, $StMYBA_1$, $StMYB_{113}$. While these three genes are involved in anthocyanin accumulation in potato tuber, it has been reported that they are all expressed positively not only in purple tubers but also in white tubers. In other words, MYB TF is not the core factor that determines anthocyanin biosynthesis in potato; instead, *StbHLH1* and *StJAF13*, encoding bHLH TFs to form the MBW complex, are required as key co-regulators (Liu et al., 2016). Another study revealed that *NtJAF13*, a bHLH homolog in *N. benthamiana*, can regulate anthocyanin biosynthesis mechanism by unilaterally affecting the transcriptional activation of *NtAN1*, encoding another bHLH TF that binds to MYB to form the MBW complex. It implies that there can be a certain inter-hierarchy between TFs involved in

anthocyanin biosynthesis (Montefiori et al., 2015; Strygina et al., 2019).

In case of eggplant, it was recently reported that some other genetic factors besides the components constituting the MBW complex can be directly related to anthocyanin biosynthesis. In eggplant, *SmMYB1* (encoding R2R3 MYB TF) and *SmTT8* (encoding bHLH TF) were known as core factors to regulate anthocyanin biosynthesis by interacting each other (Zhang et al., 2014). But the *SmWRKY44* gene, encoding a WRKY TF, was found to have significantly high expression levels in anthocyanin-rich tissues including fruit peel, and show the same expression trend as anthocyanin regulating structural and regulatory genes. Functional studies showed that the *SmWRKY44* gene was able to regulate the expression of structural genes such as *CHS*, *F3H*, and *ANS* by interacting with the promoter region of them. In addition, *SmWRKY44* was found to promote anthocyanin biosynthesis while interacting with the promoter of *SmMYB1* (He et al., 2021).

Referring to the cases of transcriptional networks in Solanaceous crops described above, it appears 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 tissues such as stems, leaves, and flowers rather than immature fruits, whereas *SlAN2-like* is specifically and strongly expressed only in immature fruits (Kiferle et al., 2015; Sun et al., 2019). However, overexpression of both *SlAN2* and *SlAN2-like* can cause anthocyanin accumulation in several tissues such as pericarps and stamens of tomato. The different characteristics between *CaAN2* and

CaAN3 identified in pepper can be inferred from the *in vivo* function of the two R2R3 MYB TFs in tomato. Phylogeny analysis showed that *SlAN2* regulating anthocyanin mainly in vegetative tissues was grouped with the fruit-specific regulator *CaAN3*, and fruit-specific regulator *SlAN2-like* was grouped with *CaAN2*, respectively (data not shown). It can be assumed that the functional differences between these transcription factor genes in tomato and pepper proceeded from divergent evolution of the progenitor of Solanaceous family (Quattrocchio et al., 1998).

Unfortunately, correlation or hierarchy between two R2R3 MYB TFs of tomato has not been studied much despite the functions of individual genes have been well characterized. Also, in tomato, the *SlMYBATV* gene encoding R3 MYB TF repressing anthocyanin biosynthesis is identified to be located on chromosome 7, and to compete with *SlAN2-like* in interacting with *SlAN1*, a bHLH TF. It will be necessary to test whether there is a homologous repressor with a similar function in pepper, and to investigate how the gene affects the expressions of *CaAN2* or *CaAN3*. In the further study, a pepper accession that has functional *CaAN2* and functional *CaAN3* but expresses only *CaAN2*, and another accession that has non-functional *CaAN2* and functional *CaAN3* thus expresses only *CaAN3* can be crossed. The F_1 plants of these two accessions should have the functional *CaAN3* allele as well as the functional *CaAN2* allele. If both genes are not expressed in F_1 plant, we can infer the relationship between the two MYB transcription factors, *CaAN2* and *CaAN3*. On the other hand, if both functional alleles are expressed, the phenotypic and genotypic segregation ratio can be calculated in an F_2 population, to infer the epistasis between

the transcription factors and find another genetic factor.

In conclusion, this study is significant since it identified a new regulatory factor involved in the regulation of anthocyanin biosynthesis in pepper and provided a hypothesis for the regulatory network. By referring to the cases of other Solanaceous crops, the study on the anthocyanin transcriptional network will continue in pepper to be helpful not only in academic sense but also in practical aspects, such as making a pepper cultivar with a high content of health functional anthocyanins.

REFERENCES

Allan AC, Espley RV (2018) MYBs Drive Novel Consumer Traits in Fruits and Vegetables. Trends Plant Sci 23:693–705.

Borovsky Y, Oren-Shamir M, Ovadia R, De Jong W, Paran I (2004) The *A* locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to *Anthocyanin2* of petunia. Theor Appl Genet 109:23–29.

Cassidy A, O'Reilly ÉJ, Kay C, Sampson L, Franz M, Forman JP, Curhan G, Rimm EB (2011) Habitual intake of flavonoid subclasses and incident hypertension in adults. Am J Clin Nutr 93:338–347.

Chepelev I, Wei G, Tang Q, Zhao K (2009) Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Res 37:1–8.

de Pascual-Teresa S, Moreno DA, García-Viguera C (2010) Flavanols and anthocyanins in cardiovascular health: A review of current evidence. Int J Mol Sci 11:1679–1703.

Dhar MK, Sharma R, Koul A, Kaul S (2015) Development of fruit color in Solanaceae: A story of two biosynthetic pathways. Brief Funct Genomics 14:199–212.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-Seq aligner.

Bioinformatics 29:15–21.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. Trends Plant Sci 15:573–581.

Gisbert C, Dumm JM, Prohens J, Vilanova S, Stommel JR (2016) A spontaneous eggplant (*Solanum melongena* L.) color mutant conditions anthocyanin-free fruit pigmentation. HortScience 51:793–798.

Gonzali S, Mazzucato A, Perata P (2009) Purple as a tomato: Towards high anthocyanin tomatoes. Trends Plant Sci 14:237–241.

He J, Monica Giusti M (2010) Anthocyanins: Natural colorants with health-promoting properties. Annu Rev Food Sci Technol 1:163–187.

He Y, Wang Z, Ge H, Liu Y, Chen H (2021) Weighted gene co-expression network analysis identifies genes related to anthocyanin biosynthesis and functional verification of hub gene *SmWRKY44*. Plant Sci 309:110935.

Hichri I, Heppel SC, Pillet J, Léon C, Czemmel S, Delrot S, Lauvergeat V, Bogs J (2010) The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. Mol Plant 3:509–523.

Jaakola L (2013) New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci. 18:477–483.

Jimenez-Garcia SN, Guevara-Gonzalez RG, Miranda-Lopez R, Feregrino-Perez AA, Torres-Pacheco I, Vazquez-Cruz MA (2013) Functional properties and quality characteristics of bioactive compounds in berries: Biochemistry, biotechnology,

and genomics. Food Res. Int. 54:1195–1207.

Jung S (2019) Identification of genetic factors controlling anthocyanin biosynthesis in *Capsicum annuum*. MS Thesis. Seoul National University. Seoul. Korea.

Jung S, Venkatesh J, Kang MY, Kwon JK, Kang BC (2019) A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in *Capsicum annuum*. Plant Sci 287:110181.

Khoo HE, Azlan A, Tang ST, Lim SM (2017) Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food Nutr Res 61:1361779.

Kiferle C, Fantini E, Bassolino L, Povero G, Spelt C, Buti S, Giuliano G, Quattrocchio F, Koes R, Perata P, Gonzali S (2015) Tomato R2R3-MYB proteins *SlANT1* and *SlAN2*: Same protein activity, different roles. PLoS One 10:1–20.

Kim J, Park M, Jeong ES, Lee JM, Choi D (2017) Harnessing anthocyanin-rich fruit: A visible reporter for tracing virus-induced gene silencing in pepper fruit. Plant Methods 13:1–10.

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: A colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242.

Lee JH, An JT, Siddique MI, Han K, Choi S, Kwon JK, Kang BC (2017) Identification and molecular genetic mapping of Chili veinal mottle virus (ChiVMV) resistance genes in pepper (*Capsicum annuum*). Mol Breed 37:1-

10.

Lin BW, Gong CC, Song HF, Cui YY (2017) Effects of anthocyanins on the prevention and treatment of cancer. Br J Pharmacol 174:1226–1243.

Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 7:1–8.

Liu Y, Lin-Wang K, Espley R V., Wang L, Yang H, Yu B, Dare A, Varkonyi-Gasic E, Wang J, Zhang J, Wang D, Allan AC (2016) Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. J Exp Bot 67:2159–2176.

Liu Y, Tikunov Y, Schouten RE, Marcelis LFM, Visser RGF, Bovy A (2018) Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. Front Chem 6:52.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-Seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517.

Mazzucato A, Willems D, Bernini R, Picarella ME, Santangelo E, Ruiu F, Tilesi F, Soressi GP (2013) Novel phenotypes related to the breeding of purple-fruited tomatoes and effect of peel extracts on human cancer cell proliferation. Plant Physiol Biochem 72:125–133.

Michelmore RW, Paran I, Kesseli R V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect

markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832.

Montefiori M, Brendolise C, Dare AP, Kui LW, Davies KM, Hellens RP, Allan AC (2015) In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. J Exp Bot 66:1427–1436.

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.

Pelletier MK, Murrell JR, Shirley BW (1997) Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis* (Further evidence for differential regulation of "early" and "late" genes). Plant Physiol 113:1437-1445.

Petroni K, Tonelli C (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci 181:219–229.

Povero G, Gonzali S, Bassolino L, Mazzucato A, Perata P (2011) Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of *Aft* and *atv* genes. J Plant Physiol 168:270–279.

Quattrocchio F, Wing JF, van der Woude K, Mol JN, Koes R (1998) Analysis of bHLH and MYB domain proteins: Species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. Plant J 13:475-488.

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3:1101–1108.

Schreiber G, Reuveni M, Evenor D, Oren-Shamir M, Ovadia R, Sapir-Mir M, Bootbool-Man A, Nahon S, Shlomo H, Chen L, Levin I (2012) *ANTHOCYANIN1* from *Solanum chilense* is more efficient in accumulating anthocyanin metabolites than its *Solanum lycopersicum* counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. Theor Appl Genet 124:295–307.

Spelt C, Quattrocchio F, Mol JNM, Koes R (2000) *Anthocyanin1* of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell 12:1619–1631.

Stommel JR, Dumm JM (2015) Coordinated regulation of biosynthetic and regulatory genes coincides with anthocyanin accumulation in developing eggplant fruit. J Amr Soc Hort Sci 140:129–135.

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J 50:660–677.

Strygina K V., Kochetov A V., Khlestkina EK (2019) Genetic control of anthocyanin pigmentation of potato tissues. BMC Genet 20:35-43.

Sun C, Deng L, Du M, Zhao J, Chen Q, Huang T, Jiang H, Li CB, Li C (2019) A transcriptional network promotes anthocyanin biosynthesis in tomato flesh. Mol Plant 13:42–58.

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi

H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R (2013) QTL-Seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183.

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. Plant J 54:733–749.

Tsukaya H, Ohshima T, Naito S, Chino M, Komeda Y (1991) Sugar-dependent expression of the *CHS-A* Gene for chalcone synthase from petunia in transgenic *Arabidopsis*. Plant Physiol 97:1414-1421.

Zhang Y, Hu Z, Chu G, Huang C, Tian S, Zhao Z, Chen G (2014) Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). J Agric Food Chem 62:2906–2912.

Zhang Z, Li DW, Jin JH, Yin YX, Zhang HX, Chai WG, Gong ZH (2015) VIGS approach reveals the modulation of anthocyanin biosynthetic genes by *CaMYB* in chili pepper leaves. Front Plant Sci 6:1–10.

ABSTRACT IN KOREAN

R2R3 MYB 전사 인자를 암호화하는 주요 조절 유전자 *CaAN2*는 고추의 다양한 조직에서 안토시아닌 생합성을 조절하는 것으로 알려져 있다. 그러나 *CaAN2*는 과실 특이적으로 안토시아닌을 발현하는 일부 고추에서 발현하지 않는다. 이 연구에서는 과일 특이적 안토시아닌 생합성을 조절하는 *CaAN3*라는 새로운 유전자좌를 확인하였다. *Capsicum annuum* 'Salad Piment Purple' 품종에서 유래한 F_2 분리집단을 구축하여 유전자 매핑을 위한 집단으로 활용하였다. F_2 집단의 식물체 과실을 샘플링하여 RNA를 추출하였고 보라색 및 녹색의 표현형으로 구분하여 풀링하였다. 각각 6개씩의 서로 다른 과실에서 추출한 RNA로 구성된 3개의 RNA 풀을 보라색, 녹색 표현형에 대해 준비하였다. Illumina 플랫폼을 통해 이 두 개의 서로 다른 RNA 풀에 대해 Bulked-segregant RNA 시퀀싱(BSR-Seq)을 수행하였다. 읽어 낸 염기서열은 고추 표준유전체인 'Dempsey'를 기준으로 정렬되었으며 두 풀 간의 Δ(SNP – index)를 계산하여 총 6,672개의 유의한 SNP를 식별하였다. 분자 표지를 통한 매핑 결과 *CaAN3*의 후보 구간을 10번 염색체 상의 184.6-186.4Mbp로 특정할 수 있었다. 이 후보 구간 안에서는 15개의 후보 유전자를 발견하였으며, 이 중 DEG 분석 결과를 기반으로 가장 강력한 후보 유전자로서 R2R3 MYB 전사 인자

인 *Dem.v1.00043895*를 선택하였다. 해당 유전자에 대한 염기서열 분석을 통해 녹색 대립 유전자의 프로모터 영역에 4개의 INDEL 변이가 있음을 확인하였다. 또한 후보 유전자의 기능 분석을 위해 virus-induced gene silencing 및 transient overexpression 분석을 수행하였다. *Dem.v1.00043895*를 침묵시켰을 때, 안토시아닌 생합성 유전자의 발현과 과피에서의 안토시아닌 축적이 감소하였다. 또한 *N. benthamiana* 잎에서 *Dem.v1.00043895*를 과발현시켰을 때 접종 부위 주변에 안토시아닌이 축적되는 것을 확인할 수 있었다. 이러한 결과를 통해 *Dem.v1.00043895*가 고추의 과실에서 안토시아닌 생합성의 조절 유전자인 *CaAN3*로 기능한다는 것을 확인하였다. 마지막으로 *CaAN3*가 일반적으로 기능하는지 여부를 테스트하기 위해 유전자 발현 분석 및 유전형 스크리닝을 24가지 다른 고추 계통에 대해 수행하였다. 과실 특이적으로 자색을 발현하는 계통들에서는 *CaAN3*만이 발현되는 것을 확인하였다. 그러나 전반적인 유전형 검토 결과 프로모터 구간의 구조적 변이가 *CaAN3*의 발현과는 직접적인 관련이 없음을 확인할 수 있었다.