Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide Association Study Reveals Candidate Genes Controlling Flavone Biosynthesis in Pepper Leaves (*Capsicum annuum* L.)

Samuel Tilahun Assefa
sumalew70@gmail.com

Ethiopian Institute of Agricultural Research
Eun-Young Yang
Korea National University of Agriculture and Fisheries
Myeong-Cheoul Cho
Korea Environmental Agriculture Institute
Jundae Lee
Jeonbuk National University

Research Article

Keywords: Capsicum Leaves Apigenin Luteolin SNP GWAS Flavonoid 3'-hydroxylase

Posted Date: September 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11002581/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Pepper (*Capsicum* spp.) leaves contain a range of flavone derivatives that occur in glycosylated, apiosylated, and/or malonylated forms. These compounds confer human health benefits and protect the plant from biotic and abiotic stress. However, to date, information on the genetic regulation of flavones in pepper leaves remains limited. Hence, a genome-wide association study (GWAS) was conducted to identify SNPs associated with pepper leaf flavone accumulation. Two multi-locus models, Blink and FarmCPU, were used in the GWAS analysis; both identified regions that may be responsible for flavone biosynthesis. The study identified several significant SNPs in linkage disequilibrium (LD) with candidate genes involved in flavonoid biosynthesis. Notably, on chromosome 3, we identified a SNP in LD with a candidate gene encoding flavonoid 3’-hydroxylase (F3’H). Cross-validation across the two mapping models (Blink and FarmCPU) indicated that F3'H could be a branch-point regulator controlling the divergence between apigenin- and luteolin-derived flavone derivatives. Other candidate genes include UGT29-family glycosyltransferases, 2OG-Fe(II) oxygenase superfamily members, and regulatory transcription factors (MYB, bHLH, NF-Y, and ERF). These findings may provide the first genome-wide genetic framework for flavone diversification in pepper leaves and a foundation for breeding efforts targeting this underexploited source of bioactive compounds.

Introduction

Pepper (*Capsicum* spp.) is one of the most widely cultivated vegetable crops worldwide. Pepper fruit is consumed both fresh and as a spice, and it is rich in polyphenols like capsaicinoids, carotenoids, vitamin C, and flavonoids(Wahyuni et al. 2011, 2013a). Polyphenols are the most abundant and structurally diverse type of plant secondary metabolites, playing important roles in plant defense, stress adaptation, and pigment formation, and providing significant health benefits in the human diet (Di Lorenzo et al., 2021; A. Khan et al., 2025; Singh et al., 2021; Zhao et al., 2022).

Flavonoids are the largest and most studied subgroup of polyphenols, with more than 9,000 flavonoids having been identified in plants (Liu et al. 2020). They protect plants from UV, biotic, and abiotic stress, and help in seed dispersal and pollination (Shi and Liu 2021; Shomali et al. 2022; Patil et al. 2024). Flavonoids are synthesized through the phenylpropanoid pathway through the action of phenylalanine lyase (PAL), cinnamic acid hydroxylase (C4H), and 4-coumarate-CoA ligase (4CL), producing the flavonoid pathway entry metabolite, *p*-coumaroyl-CoA, and the first step is committed by the action of chalcone synthase (CHS), which produces naringenin chalcone. Chalcone isomerase (CHI) converts naringenin chalcone to naringenin, from which flavones, flavonols, flavanones, anthocyanins, and isoflavones diverge (Falcone Ferreyra et al. 2012; Yonekura-Sakakibara et al. 2019a; Mao et al. 2025).

In recent years, genomic research on flavonoid biosynthesis has gained considerable interest in major crops. In pepper, for instance, early mQTL analysis of an interspecific *C. annuum* × *C. chinense* population identified 12 mQTL for flavonoid glycosides, with CaMYB12 among several co-localizing eQTL for structural pathway genes (CHS, CHI, FLS, FS-2, F3H, and F3’H) (Wahyuni et al. 2014), while a major QTL on chromosome 5 for flavonoid content was later resolved specifically to CaMYB12-like and functionally confirmed via near-isogenic lines and virus-induced gene silencing (Wu et al. 2023).

Most recently, non-targeted UHPLC-MS profiling across two mapping populations detected roughly 1,100 fruit metabolites, and combined QTL and GWAS mapping identified over 1,000 mQTL and 92 candidate genes, respectively (von Steimker et al. 2024). However, most genetic and genomic characterization of specific flavonoid branches in pepper, in both fruit and, more recently, leaf tissue, has focused almost exclusively on anthocyanin biosynthesis. SLAF-seq/BSA mapping of anthocyanin-associated regions on chromosome 10, metabolome transcriptome resolution of delphinidin-derived pigments and their MYB-related regulators, identification of a novel MBW complex (CaANT1, CaANT2, CaANI, and CaTTG1), and genome-wide DFR family characterization, identifying CaDFR5 as a key pigment-associated candidate, are examples of studies on dissecting pepper fruit anthocyanin regulation (Wang et al. 2018; Tang et al. 2020; Meng et al. 2022; An et al. 2025).

In leaf tissue, where genetic characterization of flavonoids is limited and the available studies, like fruit studies, have focused on the characterization of regulatory genes, CaMYBA/CaMYC/CaTTG1 control anthocyanin accumulation (Lu et al. 2019; Ma et al. 2025b). This pattern is consistent with a broader review of *Capsicum* phytochemical genetics, which found that gene characterization has concentrated heavily on the anthocyanin MBW complex (An2, AN3, WD40, and MYB101), leaving other flavonoid branches, including the flavone subclass, comparatively underexplored (Venkatesh et al. 2023). Flavones beyond their established antioxidant, anti-inflammatory, and anticancer activities(Desta et al. 2017; Koley et al. 2020; Akbari-Ahangar and Delnavazi 2020; Khan et al. 2021; Di Lorenzo et al. 2021), flavones hold direct breeding relevance where pepper leaves, largely overlooked relative to fruit in breeding pipelines, represent a promising alternative dietary flavone source (Cho et al. 2020; Assefa et al. 2021).

Hence, flavone-specific genetic studies in pepper leaves remain sparse and largely incidental to other research targets, in which, for instance, apigenin and apigenin-7-O-β-D-glucoside have been reported to enhance pepper's defense against *Phytophthora capsici* by inhibiting pathogen growth (Li et al. 2026). Genetic studies have remained fruit-focused, and GWAS, despite offering finer mapping resolution than QTL approaches, has rarely been applied to leaf flavonoids. Consequently, the genetic architecture underlying flavone variation in pepper leaves remains uncharacterized.

Here, we applied GWAS to 133 *Capsicum* annuum genotypes to dissect the genetic basis of the leaf flavone derivatives structurally characterized in our earlier work (Assefa et al. 2021). Cross-validation across two mapping models (Blink and FarmCPU), the study identified candidate genes such as F3'H that could be branch-point regulators controlling the divergence between apigenin- and luteolin-derived flavone derivatives, UGT29-family glycosyltransferase, 2OG-Fe(II) oxygenase superfamily members, and regulatory transcription factors (MYB, bHLH, NF-Y, and ERF).

Materials and Methods

Plant materials and DNA extraction

Pepper breeding materials representing different *Capsicum* species, i.e., *C. annuum*, *C. baccatum*, *C. frutescens*, and *C. chinense*, were obtained from the National Institute of Horticultural and Herbal Science (NIHHS), Republic of Korea. These plant materials were composed of double-haploid lines, local collections, introduced materials, hybrids, and some released varieties. Genomic DNA was extracted from two or three young pepper leaves according to a modified CTAB extraction protocol (Clarke 2009). The quantity of extracted DNA was measured using a NanoDrop (ND-1000) spectrophotometer, and its quality was assessed by the absorbance ratio (260nm/280nm > 1.8) and band integrity on a 0.5% agarose gel. Finally, the DNA concentration was adjusted to 100 ng•μL^{-1}.

Metabolite identification

In our previous paper (Assefa et al. 2021), we identified flavone derivatives from the same pepper genotypes used in this study. Hence, their relative content and genotype information were used for GWAS analysis.

Table 1
List of flavone derivatives used for GWAS analysis.

Flavones	Code
Apigenin 7-O-(2”-O-apiosyl) glucoside	AaG
Apigenin-O- (apiosyl malonyl) hexoside	AamGal
Apigenin 7-O-(2”-O-apiosyl-6”-O-malonyl) glucoside	AamG
Apigenin-O-(malonyl) hexoside	AmG
Apigenin 7-O-glucoside	AG
Luteolin-O- (apiosyl malonyl) hexoside	LamGal
Luteolin 7-O-(2”-O-apiosyl) glucoside	LaG
Luteolin 7-O-(2”-O-apiosyl-6”-O-malonyl) glucoside	LamG
Luteolin-O-(malonyl) hexoside	LmG
Luteolin 7-O-glucoside	LG
Chrysoeriol-O- (apiosyl malonyl) hexoside	ChamG

Genotyping-by-sequencing (GBS)

The GBS library was constructed from 154 DNA samples according to the method described by (Elshire et al. 2011), where single digestion with *Ape*K restriction enzyme was employed to reduce the genome complexity. The HiSeq X Ten (Illumina, San Diego, CA, USA) platform was used to sequence libraries with paired-end reads at SEEDERS (Daejeon, Republic of Korea). High-quality sequences were retained by trimming out poor reads using the DynamicTrim (phred score ≥ 20) and LengthSort programs (short read length ≥ 25bp) in the SolexaQA package(Cox et al. 2010) and they were aligned to the pepper reference genome (*Capsicum annuum* L. ver. 1.6)(http://peppergenome.snu.ac.kr/). Burrows-Wheeler Alignment (BWA) program(seed length (− l) = 30, maximum differences in the seed (− k) = 1, number of threads (− t) = 20, mismatch penalty (− M) = 6, gap open penalty (− O) = 15, and gap extension penalty (− E) = 8) and SAMtools programs were used for sequence alignment and raw SNP detection(minimal mapping quality (− Q) of 30), respectively (Li and Durbin 2009; Li et al. 2009). To confirm the accuracy and reliability of SNP genotypes, the methodology described by ((Kim et al. 2014) was used for SNP validation and to generate an integrated SNP matrix.

Population structure and linkage disequilibrium

An admixture model-based approach was implemented to estimate population structure using the STRUCTURE software ver. 2.3.4. The SNP markers were thinned using Tassel at a minimum distance of 20 kb, retaining 14,315 SNPs, and the model was run with 10,000 burn-in iterations followed by 10,000 Markov Chain Monte Carlo (MCMC) iterations. Ten replications were run for each selected K (K = 2 to K = 10). The appropriate K value was determined based on the DeltaK value estimated by Structure Harvester software using the Evanno method. (Earl and vonHoldt 2012). Principal component analysis (PCA) and linkage disequilibrium (LD) were analyzed using Tassel 5.2.96 with the default settings(Bradbury et al. 2007). The first three PCs were plotted using the ggplot package in R. LD was measured as the squared allele-frequency correlation (r^2) for all pairwise SNP comparisons within each chromosome. The LD decay curve was constructed by plotting r^2 values against the physical distance between SNPs in R (R Core Team). LD decay distance was determined at a distance when r^2 = 0.2(Nouraei et al. 2024)

Haplotype construction and haplotype trait association

Haplotype block structures around GWAS loci and candidate genes were visualized using LDBlockShow ver. 1.40 (Dong et al. 2021). Haplotypes were extracted within the LD decay window of each significant SNP, and their associations with the traits were evaluated using the geneHapR R package (Zhang et al. 2023). Only haplotypes with a frequency of 5% or more were kept for the analysis.

GWAS analysis

The Genome Association and Prediction Integrated Tool (GAPIT3) R package was employed to conduct the GWAS analyses using the two multilocus models: Bayesian-information and linkage disequilibrium iteratively nested keyway (BLINK), and Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al. 2016; Huang et al. 2019). We first transformed the relative content of the eleven flavone derivatives using rank inverse transformation and conducted the GWAS on only the 133 *C. annuum* genotypes, following population structure analysis (see results/discussion). The principal components derived from the genotype data were included as covariates to avoid spurious associations arising from population genetic structure. For all trait and marker association analyses, quantile-quantile (QQ) plots were used to assess the extent to which the models accounted for population structure and relatedness. Significant SNPs were identified using the Bonferroni correction (0.05/24764) and the Benjamini-Hochberg false discovery rate (FDR) procedure implemented in GAPIT. Candidate gene identification was performed by extracting a 600 bp sequence flanking each significant SNP (300 bp upstream and 300 bp downstream) from the CM334 reference genome. These flanking sequences were queried against the CM334 reference genome using BLASTN on the NCBI platform to confirm chromosomal coordinates and identify annotated genes within the region (https://www.ncbi.nlm.nih.gov). All genes located within the LD decay distance were considered for candidate gene analysis. For genes with hypothetical or unknown functions, the corresponding protein sequence was subjected to a BLASTP search in NCBI and the Sol Genomics Network (SNG) (https://solgenomics.net/organism/Capsicum_annuum/genome) to predict putative protein function. Gene annotations from the UCD10Xv1.1 *C. annuum* reference genome were cross-referenced to complement CM334-based annotations where applicable.

Results

Natural variation and correlation structure of selected flavone derivatives in pepper leaves

Flavone derivative concentrations in pepper leaves exhibited wide variations across the 133 pepper genotypes (Fig. 1A, Table 1). Relative concentrations were estimated using 6-methoxyluteolin as an internal standard added to each sample at 50 ppm. Among apigenin derivatives, apigenin 7-O-(2"-O-apiosyl-6"-O-malonyl) glucoside (AamG) and apigenin 7-O-(2"-O-apiosyl) glucoside (AaG) were the predominant compounds, ranging from 0.673 to 76.04 $mg \bullet g^{-1}$ DW and 0.09 to 20 $mg \bullet g^{-1}$ DW, respectively. Apigenin 7-O-glucoside (AG) and apigenin-O-(malonyl) hexoside (AmG) were present in lower quantities, ranging from 0.01 to 2 $mg \bullet g^{-1}$ DW and 0.03 to 4.13 $mg \bullet g^{-1}$ DW, respectively. Among the luteolin derivatives, luteolin 7-O-(2"-O-apiosyl-6"-O-malonyl) glucoside (LamG) was the most abundant flavone derivative, ranging from 0.3 to 62.75 $mg \bullet g^{-1}$ DW and with a mean concentration of 25 $mg \bullet g^{-1}$ DW, followed by luteolin 7-O-(2"-O-apiosyl) glucoside (LaG), ranging from 0.06 to 16.12 $mg \bullet g^{-1}$ DW. Luteolin 7-O-glucoside (LG) and luteolin-O-(malonyl) hexoside (LmG) showed mean concentrations of 1.2 and 1.6 $mg \bullet g^{-1}$ DW, respectively. The malonylated apiosyl chrysoeriol glycoside (ChamG) was the least abundant compound, ranging from 0.05 to 0.8 $mg \bullet g^{-1}$ DW with a mean of 0.3 $mg \bullet g^{-1}$ DW. The galactosidase isomers, i.e., AamGal and LamGal, had average concentrations of 1.5 and 2.2 $mg \bullet g^{-1}$ DW, respectively (Fig. 1A).

Correlation analysis of flavone derivatives revealed a clear opposing relationship between apigenin and luteolin derivatives (Fig. 1B). The strongest negative correlation was observed between AamG and LamG (r = -0.86), the two most abundant compounds consistently detected across all genotypes. Within each series, metabolites showed a strong positive correlation with other members of the same group. AG showed the highest correlation with AmG (r = 0.89), followed by AamG (r = 0.87), while the highest correlation from the luteolin group was observed between LmG and LamG (r = 0.96), followed by LG and LamG (r = 0.86). ChamG showed the highest positive correlation with LamG, while it showed a negative correlation with apigenin derivatives (Fig. 1B).

Population structure and linkage disequilibrium

A total of 1,144,529 integrated SNP matrices of 154 pepper samples, including all species, were generated, and after quality control (QC) with a minor allele frequency (MAF) > 0.05 and missing data < 30%, a total of 57,257 high-quality SNPs were retained for downstream analysis. Principal component analysis (PCA) was performed using the full set of filtered SNPs across all 154 genotypes to assess the population structure. As shown in Fig. 2A, two major clusters were observed in the PCA plot, with PC1 as the primary driver of group separation, explaining about 40% of the total genotypic variation. *C. frutescens*, *C. chinense*, and *C. baccatum* were clearly separated from the *C. annuum* genotypes along PC1.

Moreover, population structure analysis using the admixture model showed the maximum delta K value at K = 2, dividing the population into two major subpopulations (Fig. 2B and 2C), similar to the PCA result above. Hierarchical clustering analysis, based on a genetic distance matrix from the full panel, supported the PCA grouping and the admixture model with two major clusters. Genotypes of *C. frutescens*, *C. chinense*, and *C. baccatum* were consistently grouped into one cluster, while the rest of the *C. annuum* made the second cluster with several subclusters (Fig. 2E). In the full panel, about 133 (86%) genotypes were *C. annuum*, while the remaining 21 (14%) were interspecific genotypes from *C. frutescens*, *C. chinense*, and *C. baccatum*.

To evaluate population structure and LD structure within C. annuum and among the sub-populations, we refiltered SNPs at a MAF threshold of 0.05 within each subset independently using Tassel. This yielded 24,764 high-quality SNPs for the 133 *C. annuum* genotypes and 49,482 for the 21 interspecific genotypes. PCA conducted on the 133 *C. annuum* genotypes alone showed that PC1 and PC2 explained approximately 26% of the variation, substantially lower than 47% explained by PC1 and PC2 in the full-panel PCA (Fig. 2D). The marked difference in SNP contribution between the two groups, with the 21 interspecific genotypes yielding approximately twice as many high-quality SNPs as the 133 *C. annuum* genotypes, despite representing only 14% of the full panel, prompted separate linkage disequilibrium (LD) analyses within each group to assess whether this asymmetry in marker density corresponded to underlying differences in LD structure.

Comparison of LD decay curves across the three panels (full panel, *C. annuum*, and 21-genotype interspecific panel) revealed distinct patterns (Fig. 2F). The *C. annuum* panel showed a relatively slow but declining r^2 as physical distance increased, plateauing around r^2 = 0.1. In contrast, both the full panel and the 21-genotype interspecific panel showed an early plateau at higher r^2 values, approximately 0.3 in the full panel and 0.4 in the 21-genotype panel (Fig. 2F), while genome-wide LD decay distance based on the *C. annuum* genotypic information alone at r^2 = 0.2 was approximately estimated to be 120 kb (Fig. 2G).

Dual model-based GWAS analysis

GWAS was performed on 133 *C. annuum* genotypes using 24,764 high-quality SNPs. We used the multi-locus models Blink and FarmCPU to conduct GWAS, including principal components as covariates in each model (Fig. 4A and 4B). Model performance was evaluated based on the generated Q-Q plots (Fig. 4B). The Blink and FarmCPU models identified 41 and 45 significant SNPs, respectively. These SNPs were associated with nine flavone derivatives (Table 2 and Table S1), with 15 (19.7%) SNPs common to both models. Of these common SNPs, 11 tagged the same traits in both models, while the remaining four were associated with different flavones. Hence, using a single model would have missed more than 30% of the associations identified by both models. The models also identified 7 SNPs associated with at least two traits. For instance, a SNP, *CHR3_270232022*, identified by both models, was associated with six flavone derivatives, of which five were the same metabolites (both models), namely AG, AaG, AamG, AmG, and LG (Fig. 5). SNPs such as *CHR3_9067258, CHR2_48829429*, *CHR9_222240351*, and *CHR7_245315306* were associated with different metabolite types across the models. However, in this study, the trait switch between models was limited to a similar set of metabolites, i.e., apigenin and luteolin derivatives. From the Blink model, SNPs *CHR3_9067258* and *CHR2_48829429* were associated with LaG and LmG, respectively, while they were associated with LmG and LamG by the FarmCPU model. Similarly, *CHR9_222240351* and *CHR7_245315306* were associated with AaG and AmG using the Blink model, but both were associated with AG using the FarmCPU model.

Table 2
List of significant SNP variants commonly identified by Blink and FarmCPU models.

No.	SNPs	Chr	Pos	MAF	Gene ID/CM334	Annotation Description	Distance(kb)	Flavones	*P* value	Effect size	P'
1	*CHR2_11145255*	2	111145255	0.144	T459_04788	Pectate lyase	10	AmG	7.31E-09	0.493	4.
									1.3E-09	0.403	1.
2	*CHR2_164769173*	2	164769173	0.466	T459_06873	Homeobox-leucine zipper protein HOX5	21	AamG	2.73E-13	0.350	4.
					T459_06872	Bidirectional sugar transporter SWEET6a	4		1.44E-12	0.370	2.
3	*CHR3_6880678*	3	6880678	0.180	T459_07495	pulegone reductase	0.6	LaG	1.57E-10	-0.486	1!
									7.08E-09	-0.403	1⁴
4	*CHR3_270232022*	3	270232022	0.451	T459_10360	Flavonoid 3'-monooxygenase/ Flavonoid 3'-hydroxylase(F3'H)	3.2	AaG	1.57E-14	-0.426	3⁴
									1.02E-08	-0.269	2⁴
								AamG	7.59E-13	-0.333	3!
									6.77E-08	-0.259	2!
								AmG	2.38E-09	-0.400	2!
									2.41E-10	-0.349	2!
								AG	6.20E-13	-0.369	2!
									6.29E-08	-0.499	2!
								LG	5.66E-08	0.330	1!
									1.67E-11	0.392	2!
5	*CHR3_271041775*	3	271041775	0.418	T459_10410	Two-component response regulator-like APRR1	56	LmG	1.44E-19	0.496	2!
					T459_10407	Xyloglucan 6-xylosyltransferase 1	105				
					T459_10408	Xyloglucan galactosyltransferase KATAMARI1	70		5.57E-09	0.313	1!
					T459_10417	Galactoside 2-alpha-L-fucosyltransferase	6				
6	*CHR3_2737076226*	3	273707626	0.120	T459_10520	Methyl-CpG-binding domain-containing protein 11	0	LG	4.5E-10	-0.478	2!
									3.6E-07	-0.344	1!
7	*CHR6_240235334*	6	240235334	0.383	T459_18466,	"Glutathionyl-hydroquinone reductase	5	LG	2.75E-07	0.293	7.
					T459_18468	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	14		2.75E-07	0.259	7.
8	*CHR7_213079561*	7	213079561	0.364	T459_19986	Glutamate receptor 3.4	88	LaG	8.38E-10	-0.450	2!
									1.76E-15	-0.550	2!
9	*CHR7_245368893*	7	245368893	0.083	T459_20762	putative beta-1,3-galactosyltransferase 3	110	LamG	5.57E-20	-1.360	4!

No.	SNPs	Chr	Pos	MAF	Gene ID/CM334	Annotation Description	Distance(kb)	Flavones	*P* value	Effect size	P'
					T459_20766	Aconitate hydratase 3, mitochondrial	78		8.9E-14	-1.212	6
10	*CHR10_231647297*	10	231647297	0.365	T459_26561	Myricetin O-methyltransferase/Flavonoid O-methyltransferase	27	LaG	1.53E-06	-0.251	1
									6.30E-08	-0.258	1
11	*CHR11_245839116*	11	245839116	0.132	T459_28360	Putative disease resistance RPP13-like protein	32	AamG	1.88E-06	0.282	1
									2.87E-12	0.472	2

GWAS signals and candidate genes

The most prominent pleiotropic locus, *CHR3_270232022*, was associated with seven flavone derivatives, AG, AaG, AmG, AamG, AamGal, LG, and LamG, with association *p*-values ranging from 5.66×10^{-08} to 1.57×10^{-14} and phenotypic variance explained (PVE) ranging from 8.77% to 34.6% (Table 2). This SNP was located at 3 kb downstream of a candidate gene annotated as *Flavonoid 3'-monooxygenase* (NCBI) or *Flavonoid 3'-hydroxylase* (F3'H, Sol Genomic Network) (Fig. 4C). Furthermore, haplotype-trait association analysis of AG, AamG, LG, and LamG revealed significant differences between genotypes containing CG haplotypes (dominant in apigenin derivatives) and those with AA haplotypes (dominant in luteolin derivatives) (Fig. 4D and 4E, and Fig. S2). Notably, the effect size of *CHR3_270232022* was bidirectional, consistently negative for all apigenin derivatives and positive for all luteolin derivatives (Fig. S3).

A second flavonoid hydroxylation locus was identified on chromosome 11, where SNP *CHR11_262055804* detected under the FarmCPU model was significantly associated with AaG ($p = 1.55\times10^{-06}$, PVE = 17.03) and AmG ($p = 4.84\times10^{-12}$, PVE = 13.12). This SNP was located 112 kb upstream of the *flavonoid 3',5'-hydroxylase* (F3'5'H) encoding candidate gene.

On chromosome 7, two SNPs were identified in proximity to a locus encoding a 2-oxoglutarate/Fe(II)-dependent oxygenase (2OG-Fe(II) oxygenase superfamily). SNP *CHR7_245315306* was associated with AG ($p = 2.89\times10^{-10}$, PVE = 20.97) and with AmG ($p = 1.88\times10^{-08}$ and PVE = 23.3) under FarmCPU and Blink model, respectively, while *CHR7_250683709* ($p = 4.33\times10^{-07}$, PVE = 19.8, model=FarmCPU) was associated with AamG content. Independent protein BLAST searches against NCBI and the Sol Genomic Network annotated both loci as 2OG-Fe(II) oxygenase superfamily members, located 20 kb and 70 kb from *CHR7_250683709* and *CHR7_245315306*, respectively, both within the chromosome 7 LD decay distance (Table S1).

On chromosome 10, SNP CHR10_231647297 was associated with LaG content, with a candidate gene encoding Flavonoid O-methyltransferase (FOMT) located 27 kb downstream. *CHR10_231647297* was consistent across both models; it showed a negative effect size on LaG accumulation (Table 2). Additionally, a SNP on chromosome 3, located 36 kb from the 4-coumarate COA ligase (4CL) gene, showed a significant positive association with AaG.

Furthermore, we identified a candidate gene annotated as a glycosyl group transferase in the CM334 reference genome (Sol Genomic Network) within the LD decay windows of two independent GWAS signals on chromosome 3, CHR3_9086509 and CHR3_9067258, located 22 kb and 42 kb from the gene, respectively. *CHR3_9206509*, associated with LamG, was detected only by the FarmCPU model, whereas *CHR3_9187258* was associated with LG under Blink and with LamG under FarmCPU. Cross-referencing the UCD10X v1.1 C. *annuum* reference genome (Hulse-Kemp et al 2018) identified the orthologous gene as UDP-glucosyltransferase 29 (UGT29), while NCBI blastp returned beta-glucuronosyltransferase GlcAT14B. The SNP, *CHR10_231647297* (Blink and FarmCPU model) associated with LaG accumulation was co-localized with a beta-glucuronosyltransferase GlcAT14B-like candidate gene within the LD decay window(Table 2). Another SNP, *CHR10_3253886* from the Blink model, was associated with LmG and located about 10 kb from a candidate gene annotated as UDP-glucose 4,6-dehydratase, with a negative effect size on the trait. On chromosome 6, an acetyl-CoA carboxylase carboxyl transferase gene was also found 82 kb from *CHR6_744742* variants, linked with AamG (Table S1).

In addition to individual metabolite traits, we conducted GWAS on a mathematically derived composite trait, the ratio of total luteolin to total apigenin derivatives (TLAR). GWAS on TLAR identified significant associations with several variants. Notably, *CHR3_270232022* on chromosome 3, the same locus previously associated with seven individual traits, was again significantly associated with TLAR, detected consistently by both Blink ($p = 1.05\times10^{-11}$ and PVE = 22.3) and FarmCPU ($p = 1.22\times10^{-12}$ and PVE = 35.39) models (Table S1).

Regulatory and transporter candidate genes

The GWAS analysis also identified several significant hits associated with major transcription factors (TFs). On chromosome 1, MYB family transcription factor APL was in linkage disequilibrium with variant *CHR1_25407296* ($p = 3.46\times10^{-09}$), which was associated with AG content. Furthermore, a SNP, *CHR7_248188717* ($p = 9.90\times10^{-12}$), associated with AG was mapped to a genomic region containing a tandem cluster of nuclear transcription factor Y (NF-Y) subunit; four NF-Y genes were found within the LD decay window of this marker (Table 2). On chromosome 3, *CHR3_265653798* ($p = 2.77\times10^{-08}$), associated with LamG, was located 71 kb upstream of bHLH79, while bHLH81, on chromosome 1, was found at 29 kb from *CHR1_59645221* ($p = 4.03\times10^{-06}$) associated with LamG (all are under the FarmCPU model).

Furthermore, the FarmCPU model detected signals associated with ethylene-responsive transcription factors (ERFs). On chromosome 2, *CHR2_141431619* was located within a region annotated as an ERF4-like gene and was associated with AaG accumulation. On chromosome 6 (*CHR6_203711301*), associated

with both LG and LamG, was located 6.8 kb from an ethylene-responsive transcription factor (FarmCPU model only). An AP2-like ethylene-responsive transcription factor, AIL6 (AP2/ERF), was identified 53 kb from *CHR12_6606440* and was associated with variation in LaG content across the panel.

F-box/kelch-repeat protein candidate genes were also identified on three independent chromosomes. On chromosome 2, *CHR2_163200786* (Blink) was located 43 kb from an F-box/kelch-repeat protein gene and associated with AamG accumulation. Another F-box/kelch-repeat protein gene on chromosome 10 was also located 1.6 kb from CHR10_231647297 (Blink and FarmCPU) and was associated with LaG accumulation (Table S1). Three tandem clusters of F-box/kelch-repeat protein genes were located 2kb, 5kb, and 10 kb from CHR11_30783471 and were linked to the variation observed in LamG content. SNPs were also associated with candidate transport genes.

ABC transporter candidates were identified on three chromosomes. On chromosome 6, *CHR6_229378135* ($p = 6.66\times10^{-7}$), associated with LG, was located 80 kb from the ABC transporter G family member 11 gene. On chromosome 12, *CHR12_248567339* ($p = 7.9\times10^{-13}$), associated with LmG and LamG, was located within an ABC transporter B family 4 gene, 6 kb from ABC transporter B family 11, and 10 kb from ABC transporter B family 12. Additionally, on chromosome 4 (*CHR4_1074201*, $p = 2.01\times10^{-7}$), associated with LamG, was also located within the ABCB4 gene (0 kb distance) (Fig. 7 and Table S1).

Discussion

Pepper leaves as an alternative source of flavone

Flavones are increasingly becoming of interest to the scientific community due to their varied benefits to human health and plant systems. Flavones have shown broad-spectrum activity, including antioxidant, anti-inflammatory, antimicrobial, anticancer, and antidiabetic activity, driving their growing interest as functional food and nutraceutical components (Chang et al. 2015; Desta et al. 2017; Koley et al. 2020; Akbari-Ahangar and Delnavazi 2020; Khan et al. 2021; Di Lorenzo et al. 2021).

Apigenin and luteolin, the two central aglycones of flavone derivatives of the present study, have each been the subject of substantial pharmacological and agricultural research. Luteolin exhibits potent anti-cancer activity across multiple cancer cell lines and is recognized as an effective natural preservative due to its antioxidant and antimicrobial properties (Imran et al. 2019; Singh Tuli et al. 2022; Mod Razif et al. 2024; Dong et al. 2025). Both luteolin and apigenin show antidiabetic potential via α-glucosidase inhibition.

Beyond human health, apigenin and apigenin-7-O-glucosides enhance pepper's defense against *Phytophthora capsica* by inhibiting pathogen growth and reproduction, making both potential biopesticide candidates (Li et al. 2026). Moreover, CRISPR-edited wheat lines with enhanced apigenin content were found to show increased soil bacterial nitrogen fixation and yield(Tajima et al. 2025). The concentrations of flavone derivatives in the present study varied widely among the 133 pepper genotypes (Fig. 1A). The derivative concentrations observed in our results are comparable to the concentration of AG reported in parsley (13.40 $mg\bullet g^{-1}$ DW) and chamomile (53.20 $mg\bullet g^{-1}$ DW)(Hostetler et al. 2017). The dual nutraceutical and agronomic value of the flavones and genetic variation observed position pepper leaf as a promising target for flavone-directed breeding

Population structure analysis showed the presence of admixture

Before GWAS, the panels were analyzed for population structure so that it could be accounted for later in the GWAS to reduce spurious associations. The PCA, admixture models, and hierarchical clustering results together demonstrated clear population structure in the full panel, primarily attributed to interspecific divergence between *C. chinense*/*C. frutescens*/*C. baccatum* and *C. annuum*. The observed divergence is reported in other *Capsicum* diversity studies, where SNP-based genetic diversity analyses of pepper landraces and AFLP-based phylogenetic analyses have similarly grouped *C. chinense* and *C. frutescens* together, separating them from other *C. annuum* species (Wahyuni et al. 2013b; Pereira-Dias et al. 2019).

This divergence was directly reflected in the LD decay pattern, where irregular decay and an early plateau at inflated long-range r^2 values were observed in the full panel and 21 genotypes consisting of *C. chinense*/*C. frutescens*/*C. baccatum*. This is consistent with the effect of population structure on LD measures, arising from allele-frequency differences among divergent populations (Bercovich et al. 2025). The relatively smooth, slow decay observed in the *C. annuum* subset reflected the species' expected recombination history.

The disproportionately higher marker yield observed in the interspecific genotypes despite their small representation of the full panel (14%) further suggests that a substantial portion of ancestral genetic diversity was lost during the domestication bottleneck experienced by *C. annuum*, consistent with the reduced allelic diversity and elevated genome-wide LD typically reported in domesticated *C. annuum* relative to its wild and interspecific relatives(Liu et al. 2023). Hence, the lower PCA variance explained within *C. annuum* alone, the distinct LD pattern observed relative to the full panel, and the marked imbalance between representation and genetic contribution by the interspecific genotypes together indicate that these genotypes would disproportionately dominate the stratification covariates, justifying restricting the primary GWAS to the *C. annuum* panel alone.

Pepper leaf flavonoid structural candidate gene

The use of Blink and FarmCPU models substantially improved the detection power of the association analysis. Each model operates under distinct biological and statistical assumptions (Liu et al. 2016; Huang et al. 2019). The complementarity between the models was evident in the detection of a higher proportion of unique SNPs by each model, as well as of common SNPs. The use of multiple models has become standard practice in GWAS studies, and the gain in detection coverage observed here is consistent with findings reported in pepper, tomato, and rice. In pepper, for instance, six SNPs associated with fruit shape were commonly detected by the Blink and FarmCPU models (Kabange et al. 2023; Ro et al. 2024; von Steimker et al. 2025; Alatawi et al. 2026; Kayikci et al. 2026; Yogananda et al. 2026). The SNPs co-detected by both models in the present study carry evidential weight, as their identification across two independent analyses reduces the likelihood that these signals are statistical artifacts. On the other hand, the pleotropic nature of the detected SNPs, namely

CHR3_270232022, which was in LD with the *F3'H* candidate gene, could have resulted from the significant correlation observed among the flavone derivatives(Clauw et al. 2025). This was corroborated by key findings that the effect size of this SNP, *CHR3_270232022*, was bidirectional, negative for all apigenin derivatives and positive for all luteolin derivatives, with post-GWAS haplotype analysis showing a significant difference between the two haplogroups H001 and H002, mostly containing genotypes dominant in apigenin and luteolin metabolites, respectively. F3'H is a cytochrome P450-dependent monooxygenase that belongs to the CYP75 protein family, catalyzing the hydroxylation of the 3' position of flavonoids. F3'H proteins have been functionally characterized, and their substrate specificity was confirmed in various plant species, including *Brassica rapa, Triticum aestivum*, and *Oryza sativa*. For example, in *Brassica rapa*, recombinant BrF3'H catalyzed 3'-hydroxylation of apigenin to luteolin and other substrates to their corresponding 3'-hydroxylated products (Park et al. 2021); similarly, in hexaploid wheat, all functionally active CYP75B (F3'H) homologs catalyzed the conversion of apigenin to luteolin, as confirmed via microsome-based HPLC-MS/MS assays (Tajima et al. 2025). In rice, CYP75B4 was similarly shown to hydroxylate apigenin to luteolin during *in vitro* reconstitution of the tricin biosynthetic pathway(Lam et al. 2019).

Derived traits, such as percentages, ratios, or sums of related metabolites, have been used in GWAS to identify candidate genes. A recent study on blueberries used GWAS on percentage-based composite traits, including %acylated, %glycosylated anthocyanin, and total anthocyanin content, to successfully identify and validate candidate BAHD acyltransferase genes that control anthocyanin acylation (Mengist et al., 2025). Similarly, the ratio of metabolites connected by a biosynthetic pathway was associated with the locus controlling the flux between them. The locus FADS1, which encodes a fatty acid metabolism enzyme, was strongly associated (lower *p*-value) with the ratio of its direct product (20:4) to its substrate (20:3), consistent with the enzyme's biological function (Petersen et al. 2012). Our findings follow the same pattern. The luteolin/apigenin ratio (TLAR) represented the product-substrate pair flanking the F3'H-catalyzed hydroxylation step, and the ratio trait mapped to the same locus as the individual apigenin and luteolin derivatives, with the association independently detected by both Blink and FarmCPU. As with FADS1, this convergence of a biologically meaningful ratio trait with two independent models onto the responsible enzyme's locus provides strong evidence that F3'H could act as the central flux-regulating enzyme between the apigenin and luteolin branches in pepper leaves. Furthermore, a recent report showed differential expression of F3'H in pepper leaves, suggesting that this region (*CHR3_270232022-F3'H*) might be involved in the flux from apigenin to luteolin and could be considered a strong candidate for flavone biosynthesis in pepper leaves (Xu et al. 2026).

The other candidate genes identified were 2OG-Fe(II) oxygenase and well-known superfamilies that encode enzymes, including flavone synthase I (FNSI), flavanone 3-hydroxylase, and flavonol synthase (Cheng et al. 2014; Yonekura-Sakakibara et al. 2019b; Wei et al. 2021). Flavone synthase I (FNSI) has been shown to catalyze the conversion of naringenin to apigenin in mulberry and maize (H. Li et al., 2020; Righini et al., 2019). Hence, the identification of two independent 2OG-Fe(II) oxygenase candidate genes associated with apigenin derivatives suggests their potential roles in apigenin synthesis or regulation in pepper leaves. The other key candidate gene identified in our study was flavonoid O-methyltransferase (FOMT), which is associated with luteolin derivatives. FOMT catalyzes O- methylation of flavonoids, enhancing their stability and biological activity. For instance, in wheat, rice, and maize, FOMT has been reported to convert luteolin into chrysoeriol (Lam et al. 2015; Eloy et al. 2017). In this study, the chrysoeriol derivative did not show a significant association; however, the presence of FOMT near a variant associated with LAG (luteolin derivatives) may reflect a substrate-product relationship between the two metabolite classes. We have identified several chrysoeriol derivatives in these panels (Assefa et al. 2021). Hence, this region may warrant gene expression studies to validate FOMT's functional role. The above candidate genes are considered structural flavonoid genes directly involved in the biosynthesis of the flavone backbone; however, a gene found upstream in the phenylpropanoid pathway was also detected in this study. The 4-coumarate:CoA ligase (4CL) gene showed a positive association with apigenin 7-O-(2"-O-apiosyl) glucoside. 4CL catalyzes the conversion of p- coumaric acid to *p*-coumaroyl-CoA, the branch-point precursor supplying both the flavonoid and lignin biosynthetic pathways (Cao et al. 2016; Afifi et al. 2022). The positive effect size is consistent with enhanced 4CL-mediated availability of CoA substrates, which may have increased carbon flux toward downstream flavone biosynthesis, i.e., apigenin.

GWAS hit flavonoid modification genes

Flavonoid modifications are catalyzed by enzymes, including glycosyltransferase, acyltransferase, and methyltransferase. They add glycosyl, acyl, and methyl groups and are responsible for the existence of diverse flavonoid derivatives in plants (Alseekh et al. 2020). Glycosylation is one of the most prevalent modifications of flavonoids in plants, contributing to increased solubility and storage stability (Hostetler et al. 2017; Campos et al. 2019; Tang et al. 2025). This reaction is primarily catalyzed by UDP-dependent glycosyltransferase (UGTs). For instance, a CsUGT75L12, a UGT from the tea plant expressed in transgenic *Arabidopsis*, showed glucosyltransferase activity at the 7-OH position, resulting in reduced concentrations of quercetin, kaempferol, and flavan-3-ols. It was also suggested that its expression might be linked to the observed biosynthesis of luteolin-7-O-β-D-glucoside (Dai et al. 2017; Jiang et al. 2018). In sweet orange (*Citrus sinensis*), overexpression of the gene CsUGT76F1 was associated with increased levels of quercetin 7-O-rhamnoside, quercetin 7-O-glucoside, and kaempferol 7-O-glucoside (Liu et al. 2018). In our study, all flavone derivatives identified were decorated with one or two sugar moieties, reflecting the constitutive glycosylation of flavones in pepper leaves. The presence of UGT29 as a candidate gene within the LD decay window of variants associated with LG and LamG accumulation suggests that this enzyme might be involved in the glycosylation of luteolin derivatives.

In contrast, the two other detected candidate genes, GlCAT14B-like and UDP-glucose 4,6-dehydratase, might be linked to the glycosylated traits through UDP-sugar partitioning. In plants, UDP-glucose is central to multiple glycosylation pathways (Henry Christopher and Van den Ende 2018). GlCAT14B-like is a glucuronosyltransferase that uses UDP-glucuronic acid as a sugar donor to add glucuronic acid to the plant cell wall. UDP-glucuronic acid is produced from UDP-glucose by the action of UDP-glucose 4,6-dehydratase, which diverts UDP-glucose into deoxy-sugar pathways (Oka et al. 2007; Broach et al. 2012; Henry Christopher and Van den Ende 2018). Moreover, acetyl-CoA carboxylase carboxyltransferase beta subunit is another candidate gene that could participate in flavone decoration. It is a fatty acid biosynthetic enzyme that converts acetyl-CoA into malonyl-CoA(Cui et al. 2017), and its detection in our study, where malonylated flavones are the major compounds, suggested that flavone and fatty acid biosynthetic pathways in pepper leaves may compete for a shared precursor, malonyl-CoA.

The present study applies GWAS to flavone traits previously identified and structurally characterized in our earlier work, where MS fragmentations confirm the same pattern of decoration (glycosylation, malonylation, and apiosylation) across the aglycones (Assefa et al. 2021). The convergence of these structural

and metabolite correlation patterns (positive within each series, but negative between series (apigenin vs. luteolin)) with the pleiotropic F3'H region leads us to propose a tentative biosynthetic pathway in which F3'H is proposed to act as the branch-point hydroxylase directing flux toward the luteolin series at the expense of the apigenin series, with downstream decoration occurring sequentially on each aglycone scaffold. This proposed pathway is partially supported by the celery (*Apium graveolens*) biosynthetic route, in which F3'H, GlcT, and the apiosyltransferase (AgUGT94AX1) act sequentially to produce apiosylated apigenin and luteolin glucosides, which are identical to our AaG and LaG (Yoshida et al 2023). The apiosylation step in celery has no corresponding locus detected in our GWAS. Still, it is chemically preceded by the celery AgUGT94AX1 reaction, which performs AG to AaG and LG to LAG conversion in the analogous pathway. Malonylation in the pathway is inferred here solely from our own MS structural result. Given the scarcity of prior genetic work on pepper leaf flavones, this should be regarded as a well- supported starting hypothesis rather than a mechanistically validated pathway. It requires further genomic analysis, expression analysis, and functional knockout/overexpression studies

Regulatory and transporter candidate genes

In addition to these candidate genes, the GWAS analysis identified significant hits associated with major transcription factors (TFs), including MYB, bHLH, NF-Y, and ERF, which are known to regulate flavonoid biosynthetic genes. Recently, transcriptomic analysis of *Capsicum* fruits at different stages has identified co-expression of MYB with genes involved in anthocyanin, phenylpropanoid, and flavonoid biosynthesis, including cinnamate 4-hydroxylase (C4H), 4-coumaroyl-4-coumaroyl-CoA ligase (4CL), dihydroflavonol 4-reductase (DFR), flavonoid 3', 5'-hydroxylase (F3'5'H), and chalcone synthase (CHS)(Islam et al., 2021). Our GWAS detected some of these genes, such as 4CL and F3'5'H, as associated with our significant hits. Another study showed negative regulation of CaMYB108a on anthocyanin content, while its silencing resulted in upregulation of anthocyanin- and quercetin-synthesizing genes; however, genes such as flavone synthase were downregulated, resulting in reduced luteolin accumulation (He et al. 2026). Expression analysis of 122 CabHLH genes identified in the pepper genome broadly indicated their potential role in plant development and stress responses (Zhang et al. 2020). Although their study did not specifically address flavonoid metabolism, bHLH genes have been well characterized and, along with MYB and WD40, form the MBW protein complex, a well-studied regulator of flavonoid structural genes in plants. In tomato, SlJAF13 is a well-characterized bHLH factor that promotes anthocyanin biosynthesis by activating SlAN1 and forming a regulatory link with SlMYB76 to co-regulate the SlAN1 biosynthetic gene (Ma et al. 2025a).

In the case of NF-Y, there is limited information on the role of NF-Y on flavonoid biosynthesis and the presence of NF-Y genes in these GAWS hits is a particularly interesting finding in pepper that requires more detailed transcriptional and functional analysis however, a report on tomatoes showed that the suppression of a specific NF-YB gene was found to downregulate the transcription of chalcone synthase (CHS1), leading to pink colored fruit with colorless peels (Wang et al. 2021).

The ERF candidate genes detected in the present GWAS have been characterized as flavonoid regulators across a range of plant species. In crabapple, ERFs such as RAP2-4 and RAV1 have been reported to activate and repress the promoters of proanthocyanidin biosynthetic genes, respectively (Li et al. 2020a). Increased flavanol accumulation in Chardonnay grape skin was reported through vvERF003-induced upregulation of FLS1 and CHS1 (Liang et al. 2026). In tomato, SlERF7 has been shown to promote the transcriptional activity of SlPAL5, thereby inducing UV-C-induced biosynthesis of phenolic compounds (Wu et al. 2024). Hence, the ERF detected in this study, particularly the ERF4-like gene on chromosome 2, with the variant located within the gene itself, would be a region of interest for further studies to dissect the role of ERF in flavone synthesis in pepper leaves.

Unlike the transcription regulators discussed above, Kelch F-Box (KFB) proteins act post-translationally, regulating phenylpropanoid flux through targeted protein degradation. In *Arabidopsis*, a proteolytic regulator of CHS, KFB^{CHS}, was found to degrade chalcone synthase. At the same time, KFB01, KFB20, and KFB50 were shown to control the proteolytic turnover of phenylalanine ammonia lyase (PAL), directly reducing flavonoid accumulation(X. Zhang et al., 2013, 2017).In the present study, the identification of six F-box/kelch candidates at three chromosomal loci (chromosomes 2, 10, and 11) associated with AamG, LaG, and LamG could suggest that F-Box mediated proteolytic regulation of phenylpropanoid flux may contribute to the natural variation observed in flavone accumulation, prompting further functional validation studies.

Beyond biosynthetic and regulatory genes, several significant GWAS hits were associated with genes involved in vacuolar transport of flavonoids, a critical step for flavonoid sequestration and accumulation in plant cells. Flavonoids synthesized in the cytosol are transported and stored in the vacuole through mechanisms involving ATP-binding cassette (ABC) transporters, multidrug and toxic compound extrusion (MATE), and glutathione S-transferase (GST) (Petrussa et al. 2013). ABC transporters, particularly multidrug-associated proteins (MRPs), facilitate ATP-dependent vacuolar sequestration of flavonoids and their derivatives. The *Arabidopsis* ABC transporter AtAbCC2 demonstrated MgATP-dependent uptake of cyanidin 3-O glucoside into vacuolar vesicles (Behrens et al. 2019). ABCC1, a protein from grapes, when expressed in yeast cells, catalyzed the transport of malvidin 3-*O*-glucoside into the vacuole in the presence of GSH(Francisco et al. 2013). As presented earlier, three ABC transporter genes located well within the LD decay distance were identified on chromosome 12, and with a significant trait-genotype association observed at the lead SNP, this region therefore represents a strong candidate for future functional studies of flavone transport in pepper leaves.

Conclusion

Pepper leaves should be considered an underexploited source of flavones, compounds with a range of human health-promoting effects and agricultural uses. The study highlighted the importance of understanding linkage disequilibrium patterns arising from allele-frequency differences when analyzing genotype panels composed of different subpopulations or species. This will guide the subsequent downstream GWAS analysis. Independent detection of F3'H by complementary GWAS models (Blink and FarmCPU), along with a significant haplotype-trait association in this region, indicates that this is a genuine candidate driving variation in the flavone composition of our pepper leaves, directing apigenin flux toward luteolin. Detecting a pleiotropic SNP at the same locus using a metabolite-trait ratio underscores the need to functionally validate this F3H gene. ABC transporters are also identified as major contributors to

variation in flavone accumulation in pepper leaves, and the three genes clustered around the significant SNP require further functional validation. This work on flavones in pepper leaves is the first of its kind and will serve as a starting framework for dissecting flavone genetic regulation.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication:

Not applicable.

Competing interests:

The authors declare no competing interests.

Funding

This research was funded by a grant from the National Institute of Horticultural and Herbal Science, Rural Development Administration, the Republic of Korea (Project No. PJ01324302).

Author Contribution

Conceptualization, S.T.A., E.-Y.Y., M.-C.C. and J.L.; Data Curation, S.T.A.; Formal Analysis, S.T.A.; Funding Acquisition and Resource E.-Y.Y. and M.-C.C.; Methodology, S.T.A., E.-Y.Y., and J.L.; Project Administration, E.-Y.Y. and J.L.; Supervision, E.-Y.Y., M.-C.C. and J.L.; Writing—Original Draft, S.T.A.; Review and Editing, E.-Y.Y., M.-C.C. and J.L.

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. Filtered SNP data and flavone content raw data generated during the current study are available from the corresponding author upon request

References

1. Afifi OA, Tobimatsu Y, Lam PY, Martin AF, Miyamoto T, Osakabe Y, Osakabe K, Umezawa T (2022) Genome-edited rice deficient in two 4-COUMARATE:COENZYME A LIGASE genes displays diverse lignin alterations. Plant Physiol 190(4):2155–2172. https://doi.org/10.1093/plphys/kiac450
2. Akbari-Ahangar A, Delnavazi MR (2020) Flavone glycosides from the aerial parts of stachys lavandulifolia vahl. Pharm Sci 26(2):198–202. https://doi.org/10.34172/PS.2020.4
3. Alatawi I, Xiong H, Alkabkabi H, Chiwina K, Qu Y, Du R, Luo Q, Ling KS, Shi A (2026) Genome-wide study of salt tolerance in USDA tomato germplasm at seedling stage. Sci Rep 16(1). https://doi.org/10.1038/s41598-025-30759-y
4. Alseekh S, Perez de Souza L, Benina M, Fernie AR (2020) The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry, vol 174. Elsevier Ltd. https://doi.org/10.1016/j.phytochem.2020.112347
5. An Y, Li N, Zhang R, Wang S, Wang J (2025) Identification and characterization of DFR gene family and cloning of candidate genes for anthocyanin biosynthesis in pepper (Capsicum annuum L). BMC Plant Biol 25(1). https://doi.org/10.1186/s12870-025-06743-z
6. Assefa ST, Yang EY, Asamenew G, Kim HW, Cho MC, Lee J (2021) Identification of α-glucosidase inhibitors from leaf extract of pepper (Capsicum spp.) through metabolomic analysis. Metabolites 11(10). https://doi.org/10.3390/metabo11100649
7. Behrens CE, Smith KE, Iancu CV, Choe Jyong, Dean JV (2019) Transport of Anthocyanins and other Flavonoids by the Arabidopsis ATP-Binding Cassette Transporter AtABCC2. Sci Rep 9(1). https://doi.org/10.1038/s41598-018-37504-8
8. Bercovich U, Rasmussen MS, Li Z, Wiuf C, Albrechtsen A (2025) Measuring linkage disequilibrium and improvement of pruning and clumping in structured populations. Genetics 229(3). https://doi.org/10.1093/genetics/iyaf009
9. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635. https://doi.org/10.1093/bioinformatics/btm308
10. Broach B, Gu X, Bar-Peled M (2012) Biosynthesis of UDP-glucuronic acid and UDP-galacturonic acid in Bacillus cereus subsp. cytotoxis NVH 391 – 98. FEBS J 279(1):100–112. https://doi.org/10.1111/j.1742-4658.2011.08402.x
11. Campos L, López-Gresa MP, Fuertes D, Bellés JM, Rodrigo I, Lisón P (2019) Tomato glycosyltransferase Twi1 plays a role in flavonoid glycosylation and defence against virus. BMC Plant Biol 19(1):1–17. https://doi.org/10.1186/s12870-019-2063-9
12. Cao Y, Han Y, Li D, Lin Y, Cai Y (2016) Systematic analysis of the 4-coumarate:Coenzyme A ligase (4CL) related genes and expression profiling during fruit development in the Chinese pear. Genes 7(10). https://doi.org/10.3390/genes7100089
13. Chang CC, Ho SL, Lee SS (2015) Acylated glucosylflavones as α-glucosidase inhibitors from Tinospora crispa leaf. Bioorg Med Chem 23(13). https://doi.org/10.1016/j.bmc.2015.04.053

14. Cheng AX, Han XJ, Wu YF, Lou HX (2014) The function and catalysis of 2-oxoglutarate-dependent oxygenases involved in plant flavonoid biosynthesis. Int J Mol Sci 15(1):1080–1095. https://doi.org/10.3390/ijms15011080
15. Cho SY, Kim HW, Lee MK, Kim HJ, Kim JB, Choe JS, Lee YM, Jang HH (2020) Antioxidant and anti-inflammatory activities in relation to the flavonoids composition of pepper (Capsicum annuum L). Antioxidants 9(10):1–11. https://doi.org/10.3390/antiox9100986
16. Clarke JD (2009) Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harbor Protoc 4(3):5177–5179. https://doi.org/10.1101/pdb.prot5177
17. Clauw P, Ellis TJ, Liu HJ, Sasaki E (2025) Beyond the Standard GWAS - A Guide for Plant Biologists. In *Plant and Cell Physiology* (Vol. 66, Number 4, pp. 431–443). Oxford University Press. https://doi.org/10.1093/pcp/pcae079
18. Cox MP, Peterson DA, Biggs PJ (2010) *SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data*.
19. Cui Y, Zhao Y, Wang Y, Liu Z, Ijaz B, Huang Y, Hua J (2017) Genome-wide identification and expression analysis of the biotin carboxyl carrier subunits of heteromeric acetyl-coa carboxylase in Gossypium. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00624
20. Dai X, Zhuang J, Wu Y, Wang P, Zhao G, Liu Y, Jiang X, Gao L, Xia T (2017) Identification of a Flavonoid Glucosyltransferase Involved in 7-OH Site Glycosylation in Tea plants (Camellia sinensis). Sci Rep 7(1). https://doi.org/10.1038/s41598-017-06453-z
21. Desta KT, Kim GS, El-Aty AMA, Raha S, Kim MB, Jeong JH, Warda M, Hacımüftüoğlu A, Shin HC, Shim JH, Shin SC (2017) Flavone polyphenols dominate in Thymus schimperi Ronniger: LC–ESI–MS/MS characterization and study of anti-proliferative effects of plant extract on AGS and HepG2 cancer cells. J Chromatogr B: Anal Technol Biomedical Life Sci 1053:1–8. https://doi.org/10.1016/j.jchromb.2017.03.035
22. Di Lorenzo C, Colombo F, Biella S, Stockley C, Restani P (2021) Polyphenols and human health: The role of bioavailability. In *Nutrients* (Vol. 13, Number 1, pp. 1–30). MDPI AG. https://doi.org/10.3390/nu13010273
23. Dong SS, He WM, Ji JJ, Zhang C, Guo Y, Yang TL (2021) LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Brief Bioinform 22(4). https://doi.org/10.1093/bib/bbaa227
24. Dong W, Wang X, Xiang M, Chen J, Zeng J, Chen M (2025) Effects of Luteolin Treatment on Postharvest Quality and Antioxidant Capacity of Nanfeng Tangerines. Foods 14(1). https://doi.org/10.3390/foods14010068
25. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. https://doi.org/10.1007/s12686-011-9548-7
26. Eloy NB, Voorend W, Lan W, Saleme M, de Cesarino LS, Vanholme I, Smith R, Goeminne RA, Pallidis G, Morreel A, Nicomedes K, Ralph J, J., Boerjan W (2017) Silencing CHALCONE SYNTHASE in maize impedes the incorporation of tricin into lignin and increases lignin content. Plant Physiol 173(2):998–1016. https://doi.org/10.1104/pp.16.01108
27. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):1–10. https://doi.org/10.1371/journal.pone.0019379
28. Falcone Ferreyra ML, Rius SP, Casati P (2012) Flavonoids: Biosynthesis, biological functions, and biotechnological applications. Front Plant Sci 3(SEP):1–15. https://doi.org/10.3389/fpls.2012.00222
29. Francisco RM, Regalado A, Ageorges A, Burla BJ, Bassin B, Eisenach C, Zarrouk O, Vialet S, Marlin T, Chaves MM, Martinoia E, Nagy R (2013) ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-glucosides. Plant Cell 25(5):1840–1854. https://doi.org/10.1105/tpc.112.102152
30. He Q, Liu S, Wang C, Liu Z, Xu S, Li X, Yu H, Yang X, Li Z, Liu W, Li W, Deng Y, Hou C (2026) The transcription factor CaMYB108a negatively regulates anthocyanin and other flavonoid biosynthesis in chili pepper (Capsicum annuum L.) fruit. *Plant Physiology and Biochemistry*, *234*. https://doi.org/10.1016/j.plaphy.2026.111305
31. van Henry Christopher J R., Van den Ende W (2018) UDP-glucose: A potential signaling molecule in plants? *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.02230
32. Hostetler GL, Ralston RA, Schwartz SJ (2017) Flavones: Food sources, bioavailability, metabolism, and bioactivity. In *Advances in Nutrition* (Vol. 8, Number 3, pp. 423–435). American Society for Nutrition. https://doi.org/10.3945/an.116.012948
33. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019) BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8(2):1–12. https://doi.org/10.1093/gigascience/giy154
34. Imran M, Rauf A, Abu-Izneid T, Nadeem M, Shariati MA, Khan IA, Imran A, Orhan IE, Rizwan M, Atif M, Gondal TA, Mubarak MS (2019) Luteolin, a flavonoid, as an anticancer agent: A review. In *Biomedicine and Pharmacotherapy*. Elsevier Masson SAS 112. https://doi.org/10.1016/j.biopha.2019.108612
35. Islam K, Rawoof A, Ahmad I, Dubey M, Momo J, Ramchiary N (2021) Capsicum chinense MYB Transcription Factor Genes: Identification, Expression Analysis, and Their Conservation and Diversification With Other Solanaceae Genomes. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.721265
36. Jiang X, Shi Y, Dai X, Zhuang J, Fu Z, Zhao X, Liu Y, Gao L, Xia T (2018) Four flavonoid glycosyltransferases present in tea overexpressed in model plants Arabidopsis thaliana and Nicotiana tabacum for functional identification. J Chromatogr B: Anal Technol Biomedical Life Sci 1100–1101:148–157. https://doi.org/10.1016/j.jchromb.2018.09.033
37. Kabange NR, Alibu S, Kwon Y, Lee SM, Oh KW, Lee JH (2023) Genome-wide association study (GWAS) with high-throughput SNP chip DNA markers identified novel genetic factors for mesocotyl elongation and seedling emergence in rice (Oryza sativa L.) using multiple GAPIT models. *Frontiers in Genetics*, *14*. https://doi.org/10.3389/fgene.2023.1282620
38. Kayikci HC, Aydin S, Adak A, Dogan A, Sapkota M, Feng Q, Topcu Y (2026) Association mapping of tomato fruit quality for weight, firmness, brix, and color using GWAS. BMC Plant Biol 26(1). https://doi.org/10.1186/s12870-025-07838-3

39. Khan A, Kanwal F, Ullah S, Fahad M, Tariq L, Altaf MT, Riaz A, Zhang G (2025) Plant Secondary Metabolites—Central Regulators Against Abiotic and Biotic Stresses. In *Metabolites* (Vol. 15, Number 4). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/metabo15040276
40. Khan AU, Dagur HS, Khan M, Malik N, Alam M, Mushtaque M (2021) Therapeutic role of flavonoids and flavones in cancer prevention: Current trends and future perspectives. Eur J Med Chem Rep 3. https://doi.org/10.1016/j.ejmcr.2021.100010
41. Kim JE, Oh SK, Lee JH, Lee BM, Jo SH (2014) Genome-wide SNP calling using next generation sequencing data in tomato. Mol Cells 37(1):36–42. https://doi.org/10.14348/molcells.2014.2241
42. Koley TK, Khan Z, Oulkar D, Singh BK, Maurya A, Singh B, Banerjee K (2020) High resolution LC-MS characterization of phenolic compounds and the evaluation of antioxidant properties of a tropical purple radish genotype. Arab J Chem 13(1):1355–1366. https://doi.org/10.1016/j.arabjc.2017.11.007
43. Lam PY, Liu H, Lo C (2015) Completion of tricin biosynthesis pathway in rice: Cytochrome P450 75B4 is a unique chrysoeriol 5'-hydroxylase. Plant Physiol 168(4):1527–1536. https://doi.org/10.1104/pp.15.00566
44. Lam PY, Lui ACW, Yamamura M, Wang L, Takeda Y, Suzuki S, Liu H, Zhu FY, Chen MX, Zhang J, Umezawa T, Tobimatsu Y, Lo C (2019) Recruitment of specific flavonoid B-ring hydroxylases for two independent biosynthesis pathways of flavone-derived metabolites in grasses. New Phytol 223(1):204–219. https://doi.org/10.1111/nph.15795
45. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324
46. Li H, Han M, Yu L, Wang S, Zhang J, Tian J, Yao Y (2020) Transcriptome Analysis Identifies Two Ethylene Response Factors That Regulate Proanthocyanidin Biosynthesis During Malus Crabapple Fruit Development. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00076
47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352
48. Li H, Li D, Yang Z, Zeng Q, Luo Y, He N (2020) Flavones produced by mulberry flavone synthase type i constitute a defense line against the ultraviolet-b stress. Plants 9(2). https://doi.org/10.3390/plants9020215
49. Li L, Zhang K, Lai B, Qiu X, Ren J, Lin L, Shi R, Liu W, Sedibe M, Bao J, Norvienyeku J (2026) Apigenin and apigenin-7-O-β-d-glucoside enhance pepper defense against Phytophthora capsici by inhibiting pathogen growth and reproduction. Phytopathol Res 8(1). https://doi.org/10.1186/s42483-026-00414-z
50. Liang J, Wang M, Wu Y, Yang C, Shang H, Zhang P, Wen P (2026) Ethylene-Mediated VvERF003 Promotes Flavonol Accumulation by Upregulating VvFLS1 and VvCHI1 in 'Chardonnay' Grape Berry Skin. Biomolecules 16(1). https://doi.org/10.3390/biom16010069
51. Liu F, Zhao J, Sun H, Xiong C, Sun X, Wang X, Wang Z, Jarret R, Wang J, Tang B, Xu H, Hu B, Suo H, Yang B, Ou L, Li X, Zhou S, Yang S, Liu Z, Zou X (2023) Genomes of cultivated and wild Capsicum species provide insights into pepper domestication and population differentiation. Nat Commun 14(1). https://doi.org/10.1038/s41467-023-41251-4
52. Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12(2). https://doi.org/10.1371/journal.pgen.1005767
53. Liu X, Lin C, Ma X, Tan Y, Wang J, Zeng M (2018) Functional characterization of a flavonoid glycosyltransferase in sweet orange (Citrus sinensis). Front Plant Sci 9(February):1–14. https://doi.org/10.3389/fpls.2018.00166
54. Liu Y, Lv J, Liu Z, Wang J, Yang B, Chen W, Ou L, Dai X, Zhang Z, Zou X (2020) Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (Capsicum annuum L.). *Food Chemistry*, *306*. https://doi.org/10.1016/j.foodchem.2019.125629
55. Lu BY, Cheng GX, Zhang Z, Sun JT, Ali M, Jia QL, Luo DX, Gong ZH, Li DW (2019) CaMYC, A Novel Transcription Factor, Regulates Anthocyanin Biosynthesis in Color-leaved Pepper (Capsicum annuum L). J Plant Growth Regul 38(2):574–585. https://doi.org/10.1007/s00344-018-9871-2
56. Ma S, Duan Z, Yin X, Chen L (2025) SlMYB76, an SlANS-Repressing R2R3-MYB Transcription Factor, Regulates Anthocyanin Accumulation in 'Black Pearl' Tomato (Solanum lycopersicum). Genes 16(11). https://doi.org/10.3390/genes16111291
57. Ma X, Liang G, Xu Z, Lin C, Zhu B (2025) CaMYBA–CaMYC–CaTTG1 complex activates the transcription of anthocyanin synthesis structural genes and regulates anthocyanin accumulation in pepper (Capsicum annuum L.) leaves. *Frontiers in Plant Science*, *16*. https://doi.org/10.3389/fpls.2025.1538607
58. Mao Y, Luo J, Cai Z (2025) Biosynthesis and Regulatory Mechanisms of Plant Flavonoids: A Review. In *Plants* (Vol. 14, Number 12). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/plants14121847
59. Meng Y, Zhang H, Fan Y, Yan L (2022) Anthocyanins accumulation analysis of correlated genes by metabolome and transcriptome in green and purple peppers (Capsicum annuum). BMC Plant Biol 22(1). https://doi.org/10.1186/s12870-022-03746-y
60. Mod Razif MRF, Chan SY, Chew YL, Hassan M, Ahmad Hisham S, Abdul Rahman S, Mai CW, Teo MYM, Kee PE, Khoo KS, Lee SK, Liew K, Bin (2024) Recent Developments in Luteolin-Loaded Nanoformulations for Enhanced Anti-Carcinogenic Activities: Insights from In Vitro and In Vivo Studies. In *Sci* (Vol. 6, Number 4). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/sci6040068
61. Nouraei S, Mia MS, Liu H, Turner NC, Yan G (2024) Genome-wide association study of drought tolerance in wheat (Triticum aestivum L.) identifies SNP markers and candidate genes. Mol Genet Genomics 299(1). https://doi.org/10.1007/s00438-024-02104-x
62. Oka T, Nemoto T, Jigami Y (2007) Functional analysis of Arabidopsis thaliana RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose Conversion. J Biol Chem 282(8):5389–5403. https://doi.org/10.1074/jbc.M610196200
63. Park S, Lee H, Min MK, Ha J, Song J, Lim CJ, Oh J, Lee SB, Lee JY, Kim BG (2021) Functional Characterization of BrF3'H, Which Determines the Typical Flavonoid Profile of Purple Chinese Cabbage. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.793589
64. Patil JR, Mhatre KJ, Yadav K, Yadav LS, Srivastava S, Nikalje GC (2024) Flavonoids in plant-environment interactions and stress responses. Discover Plants 1(1). https://doi.org/10.1007/s44372-024-00063-6

65. Pereira-Dias L, Vilanova S, Fita A, Prohens J, Rodríguez-Burruezo A (2019) Genetic diversity, population structure, and relationships in a collection of pepper (Capsicum spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic Res 6(1). https://doi.org/10.1038/s41438-019-0132-8
66. Petersen AK, Krumsiek J, Wägele B, Theis FJ, Wichmann HE, Gieger C, Suhre K (2012) On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. BMC Bioinformatics 13(1). https://doi.org/10.1186/1471-2105-13-120
67. Petrussa E, Braidot E, Zancani M, Peresson C, Bertolini A, Patui S, Vianello A (2013) Plant flavonoids-biosynthesis, transport and involvement in stress responses. In *International Journal of Molecular Sciences* (Vol. 14, Number 7, pp. 14950–14973). https://doi.org/10.3390/ijms140714950
68. Righini S, Rodriguez EJ, Berosich C, Grotewold E, Casati P, Falcone Ferreyra ML (2019) Apigenin produced by maize flavone synthase I and II protects plants against UV-B-induced damage. Plant Cell Environ 42(2):495–508. https://doi.org/10.1111/pce.13428
69. Ro N, Oh H, Ko HC, Yi J, Na YW, Haile M (2024) Exploring Genomic Regions Associated with Fruit Traits in Pepper: Insights from Multiple GWAS Models. Int J Mol Sci 25(21). https://doi.org/10.3390/ijms252111836
70. Shi C, Liu H (2021) How plants protect themselves from ultraviolet-B radiation stress. Plant Physiol 187(3):1096–1103. https://doi.org/10.1093/plphys/kiab245
71. Shomali A, Das S, Arif N, Sarraf M, Zahra N, Yadav V, Aliniaeifard S, Chauhan DK, Hasanuzzaman M (2022) Diverse Physiological Roles of Flavonoids in Plant Environmental Stress Responses and Tolerance. In *Plants* (Vol. 11, Number 22). MDPI. https://doi.org/10.3390/plants11223158
72. Singh S, Kaur I, Kariyat R (2021) The multifunctional roles of polyphenols in plant-herbivore interactions. In *International Journal of Molecular Sciences* (Vol. 22, Number 3, pp. 1–19). MDPI AG. https://doi.org/10.3390/ijms22031442
73. Singh Tuli H, Rath P, Chauhan A, Sak K, Aggarwal D, Choudhary R, Sharma U, Vashishth K, Sharma S, Kumar M, Yadav V, Singh T, Yerer MB, Haque S (2022) Luteolin, a Potent Anticancer Compound: From Chemistry to Cellular Interactions and Synergetic Perspectives. In *Cancers* (Vol. 14, Number 21). MDPI. https://doi.org/10.3390/cancers14215373
74. Tajima H, Yadav A, Castellanos JH, Yan D, Brookbank BP, Nambara E, Blumwald E (2025) Increased Apigenin in DNA-Edited Hexaploid Wheat Promoted Soil Bacterial Nitrogen Fixation and Improved Grain Yield Under Limiting Nitrogen Fertiliser. Plant Biotechnol J 23(11):5146–5160. https://doi.org/10.1111/pbi.70289
75. Tang B, Li L, Hu Z, Chen Y, Tan T, Jia Y, Xie Q, Chen G (2020) Anthocyanin accumulation and transcriptional regulation of anthocyanin biosynthesis in purple pepper. J Agric Food Chem 68(43):12152–12163. https://doi.org/10.1021/acs.jafc.0c02460
76. Tang S, Wang B, Liu X, Xi W, Yue Y, Tan X, Bai J, Huang L (2025) Structural insights and biological activities of flavonoids: Implications for novel applications. In *Food Frontiers* (Vol. 6, Number 1, pp. 218–247). John Wiley and Sons Inc. https://doi.org/10.1002/fft2.494
77. Venkatesh J, Lee SY, Back S, Kim TG, Kim GW, Kim JM, Kwon JK, Kang BC (2023) Update on the genetic and molecular regulation of the biosynthetic pathways underlying pepper fruit color and pungency. In *Current Plant Biology* (Vols. 35–36). Elsevier B.V. https://doi.org/10.1016/j.cpb.2023.100303
78. von Steimker J, Tripodi P, Wendenburg R, Tringovska I, Nankar AN, Stoeva V, Pasev G, Klemmer A, Todorova V, Bulut M, Tikunov Y, Bovy A, Gechev T, Kostova D, Fernie AR, Alseekh S (2024) The genetic architecture of the pepper metabolome and the biosynthesis of its signature capsianoside metabolites. Curr Biol 34(18):4209–4223e3. https://doi.org/10.1016/j.cub.2024.07.098
79. von Steimker J, Wendenburg R, Klemmer A, Rosaria M, Fernie AR, Alseekh S, Tripodi P (2025) Genome-wide association analysis and linkage mapping decipher the genetic control of primary metabolites and quality traits in Capsicum. Plant J 122(6). https://doi.org/10.1111/tpj.70300
80. Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG (2011) Metabolite biodiversity in pepper (Capsicum) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. Phytochemistry 72(11–12):1358–1370. https://doi.org/10.1016/j.phytochem.2011.03.016
81. Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG (2013) Secondary metabolites of Capsicum species and their importance in the human diet. J Nat Prod 76(4):783–793. https://doi.org/10.1021/np300898z
82. Wahyuni Y, Ballester AR, Tikunov Y, de Vos RCH, Pelgrom KTB, Maharijaya A, Sudarmonowati E, Bino RJ, Bovy AG (2013) Metabolomics and molecular marker analysis to explore pepper (Capsicum sp.) biodiversity. Metabolomics 9(1):130–144. https://doi.org/10.1007/s11306-012-0432-6
83. Wahyuni Y, Stahl-Hermes V, Ballester AR, de Vos RCH, Voorrips RE, Maharijaya A, Molthoff J, Zamora MV, Sudarmonowati E, Arisi ACM, Bino RJ, Bovy AG (2014) Genetic mapping of semi-polar metabolites in pepper fruits (Capsicum sp.): Towards unravelling the molecular regulation of flavonoid quantitative trait loci. Mol Breeding 33(3):503–518. https://doi.org/10.1007/s11032-013-9967-0
84. Wang G, Chen B, Du H, Zhang F, Zhang H, Wang Y, He H, Geng S, Zhang X (2018) Genetic mapping of anthocyanin accumulation-related genes in pepper fruits using a combination of SLAF-seq and BSA. PLoS ONE 13(9). https://doi.org/10.1371/journal.pone.0204690
85. Wang J, Li G, Li C, Zhang C, Cui L, Ai G, Wang X, Zheng F, Zhang D, Larkin RM, Ye Z, Zhang J (2021) NF-Y plays essential roles in flavonoid biosynthesis by modulating histone modifications in tomato. New Phytol 229(6):3237–3252. https://doi.org/10.1111/nph.17112
86. Wei S, Zhang W, Fu R, Zhang Y (2021) Genome-wide characterization of 2-oxoglutarate and Fe(II)-dependent dioxygenase family genes in tomato during growth cycle and their roles in metabolism. BMC Genomics 22(1):1–14. https://doi.org/10.1186/s12864-021-07434-3
87. Wu C, Men C, Yan L, Zhang J, Wang Y, Chen M, Liu C, Zheng L (2024) The ERF transcription factor SlERF7 promotes UV-C-induced biosynthesis of phenolic compounds in tomato. *Scientia Horticulturae*, *338*. https://doi.org/10.1016/j.scienta.2024.113643
88. Wu Y, Popovsky-Sarid S, Tikunov Y, Borovsky Y, Baruch K, Visser RGF, Paran I, Bovy A (2023) CaMYB12-like underlies a major QTL for flavonoid content in pepper (Capsicum annuum) fruit. New Phytol 237(6):2255–2267. https://doi.org/10.1111/nph.18693
89. Xu Z, Gu Y, Ma X, Yan Y, Gao Y, Lin C, Luo P, Zhu B (2026) Integrated metabolites and transcripts profiling reveal biosynthetic basis and regulation of color formation in pepper (Capsicum annuum L.) leaf. *Plant Physiology and Biochemistry*, *235*. https://doi.org/10.1016/j.plaphy.2026.111391

90. Yogananda M, Sharma VK, Srivastava A, Shashidhar BR, Saini N, Upadhyay P, Aruna TS, Kumar S, Iquebal MA, Jaiswal S, Danakumara T, Mangal M (2026) Unravelling the genetics of heat-responsive fruit traits in pepper [Capsicum annuum L.] via GWAS under subtropical climates. BMC Plant Biol 26(1). https://doi.org/10.1186/s12870-026-08379-z
91. Yonekura-Sakakibara K, Higashi Y, Nakabayashi R (2019a) The Origin and Evolution of Plant Flavonoid Metabolism. Front Plant Sci 10(August):1–16. https://doi.org/10.3389/fpls.2019.00943
92. Yonekura-Sakakibara K, Higashi Y, Nakabayashi R (2019b) The Origin and Evolution of Plant Flavonoid Metabolism. Front Plant Sci 10(August):1–16. https://doi.org/10.3389/fpls.2019.00943
93. Zhang R, Jia G, Diao X (2023) geneHapR: an R package for gene haplotypic statistics and visualization. BMC Bioinformatics 24(1). https://doi.org/10.1186/s12859-023-05318-9
94. Zhang X, Abrahan C, Colquhoun TA, Liu CJ (2017) A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in arabidopsis. Plant Cell 29(5):1157–1174. https://doi.org/10.1105/tpc.16.00855
95. Zhang X, Gou M, Liu CJ (2013) Arabidopsis kelch repeat F-Box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. Plant Cell 25(12):4994–5010. https://doi.org/10.1105/tpc.113.119644
96. Zhang Z, Chen J, Liang C, Liu F, Hou X, Zou X (2020) Genome-Wide Identification and Characterization of the bHLH Transcription Factor Family in Pepper (Capsicum annuum L.). *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.570156
97. Zhao X, Zhang Y, Long T, Wang S, Yang J (2022) Regulation Mechanism of Plant Pigments Biosynthesis: Anthocyanins, Carotenoids, and Betalains. In *Metabolites* (Vol. 12, Number 9). MDPI. https://doi.org/10.3390/metabo12090871

Figures

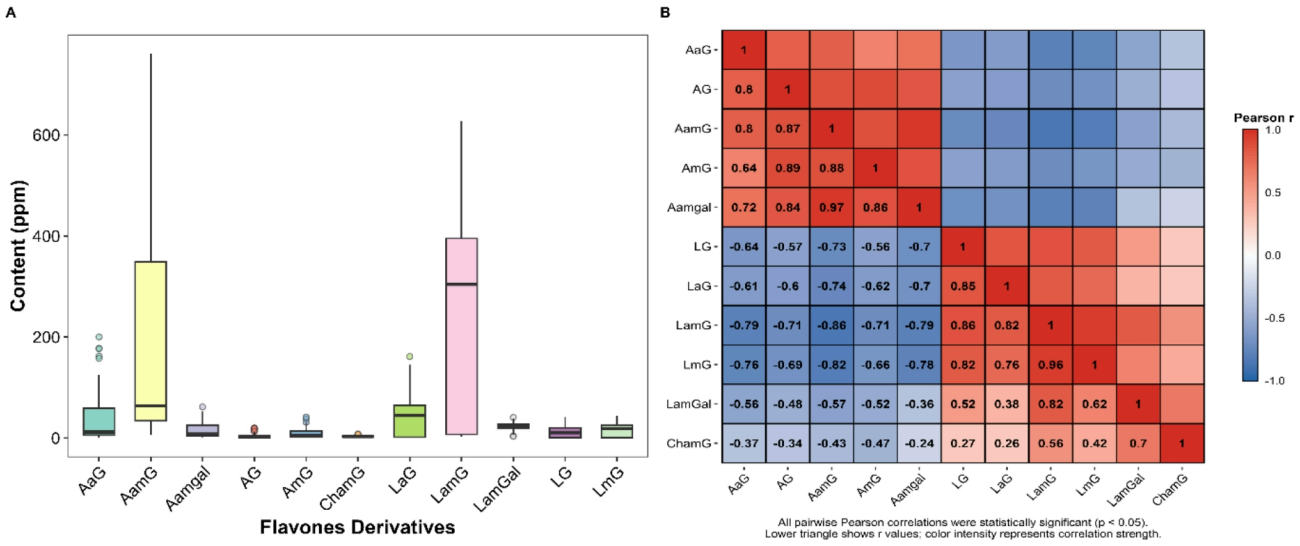


Figure 1

Pepper leaf flavone content variation (A) and their correlation (B).

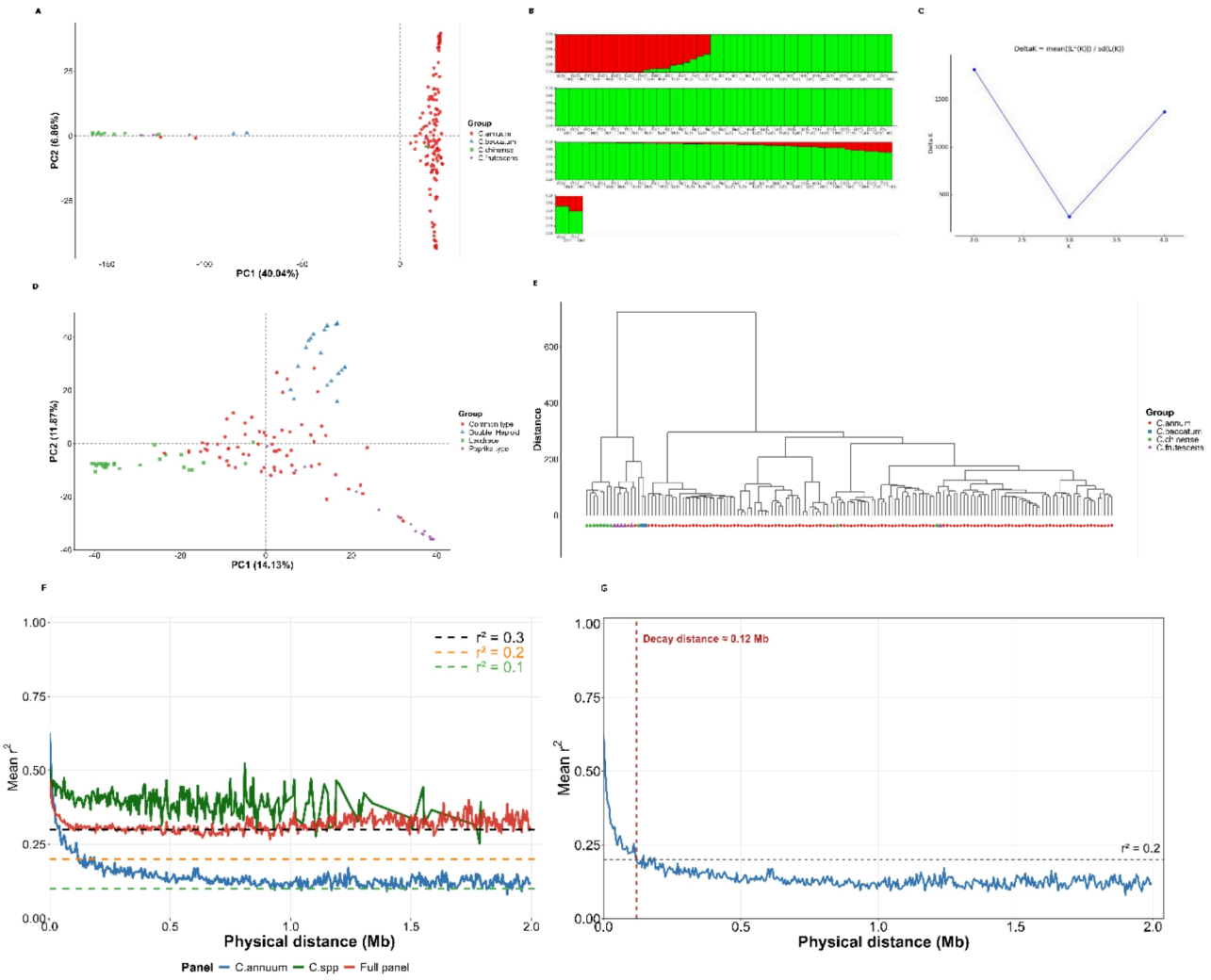


Figure 2

Population structure of the full panel depicted using PCA (A), admixture subpopulation structure shown using optimal K values (B and C), hierarchical clustering (D), and population structure of C. annuum-only genotypes using PCA (E). and Linkage disequilibrium pattern comparison among the full panel, C. annuum only, and other Capsicum spp. (C. frutescens, C. chinense, and C. baccatum) (F) and genome-wide LD decay distance estimate on the C. annuum only genotype (G).

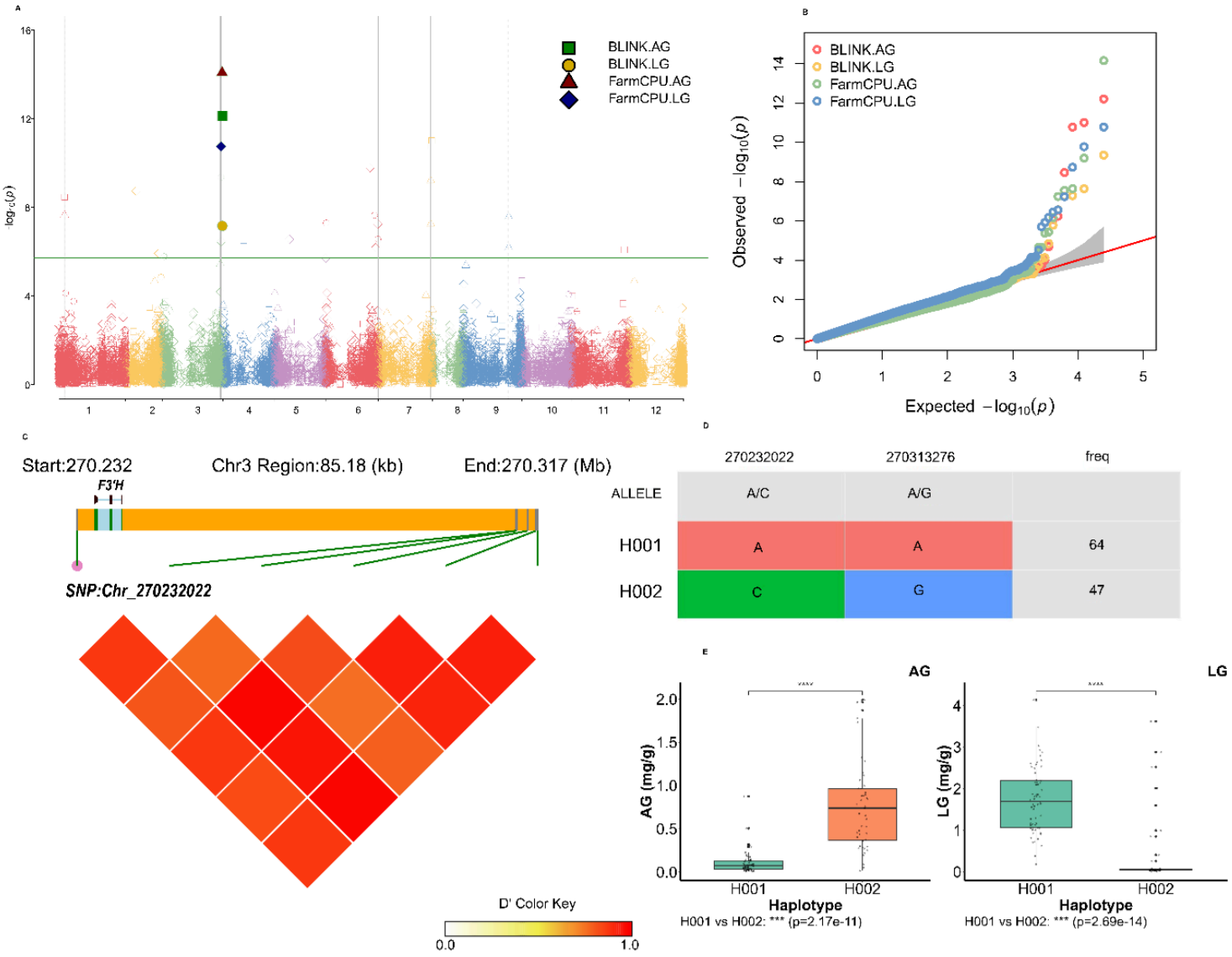


Figure 3

Pleiotropic SNP chr3_270232022 detected by both models associated with AG and LG displayed in the Manhattan plot (A), and their QQ plot exhibiting how the models fit well (B), LDBlockShow-generated LD block showing LD strength between SNPs flanking candidate gene F3'H (C), haplotype generated between chr3_270232022 and chr3_270313276 flanking the gene(F3'H) (D), and their significant association shown in the boxplot (E).



Figure 4

SNP chr12_248567339 associated with LamG displayed in the Manhattan plot (A), and SNP genotype frequency at the locus (B), three tandemly arranged ABC transporter B genes that are in LD with the SNP (C), and boxplot showing significant trait-genotype e associations (D).

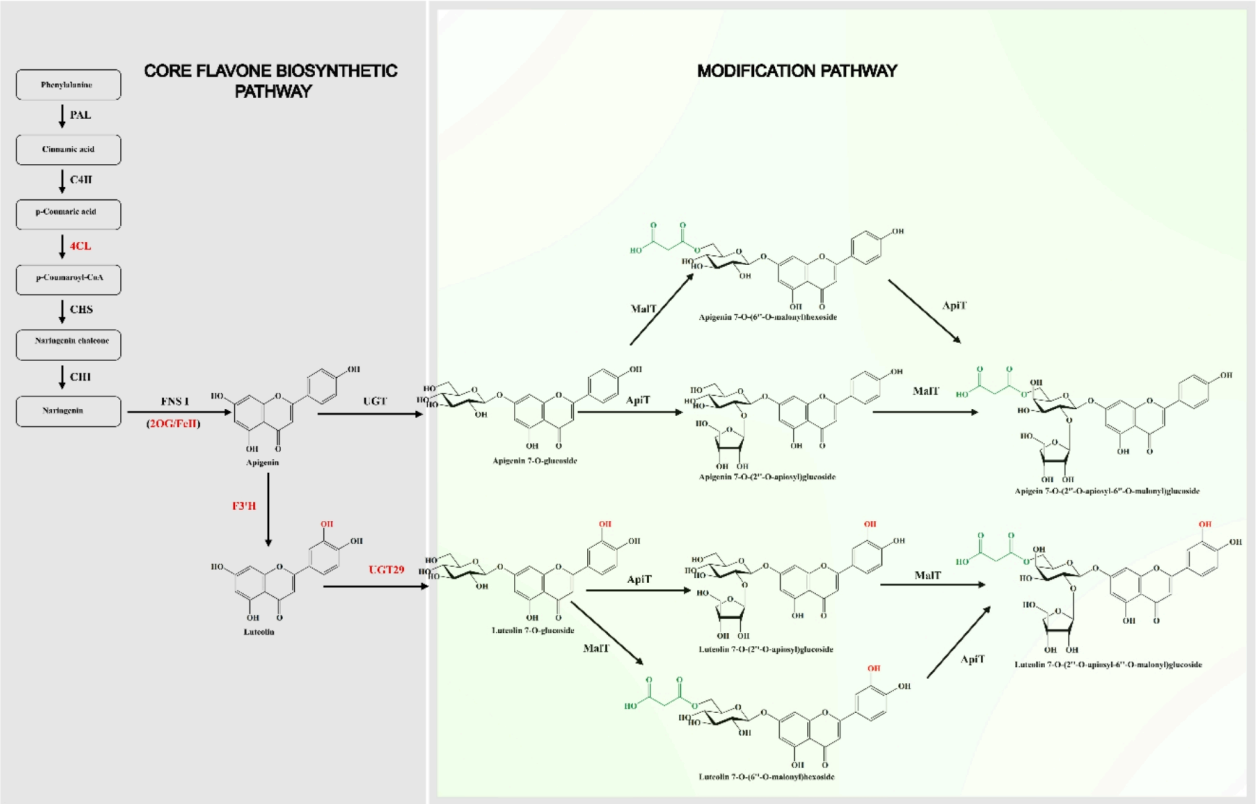


Figure 5

Proposed flavone biosynthetic pathway in pepper leaf (A) and chromatogram pattern among genotypes that are majorly apigenin-dominant. Genes highlighted in Red are candidate genes identified in this study**. ApiT**: Apiosyltransferase; **MalT**: Malonyltransferase

Supplementary Files

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

- TableS1.xlsx
- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf