Theoretical and Applied Genetics (2024) 137:130
https://doi.org/10.1007/s00122-024-04635-8

ORIGINAL ARTICLE

Genetic tapestry of *Capsicum* fruit colors: a comparative analysis of four cultivated species

Ambika Bhattarai[1] · Padma Nimmakayala[1] · Brittany Davenport[1] · Purushothaman Natarajan[1] · Krittika Tonapi[1] · Sai Satish Kadiyala[1] · Carlos Lopez-Ortiz[1] · Lizbeth Ibarra-Muñoz[1,2] · Manohar Chakrabarti[3] · Vagner Benedito[4] · Donald A. Adjeroh[5] · Nagamani Balagurusamy[2] · Umesh K. Reddy[1]

Received: 7 November 2023 / Accepted: 17 February 2024 / Published online: 14 May 2024

Abstract

Key message **Genome-wide association study of color spaces across the four cultivated *Capsicum* spp. revealed a shared set of genes influencing fruit color, suggesting mechanisms and pathways across *Capsicum* species are conserved during the speciation. Notably, Cytochrome P450 of the carotenoid pathway, MYB transcription factor, and pentatricopeptide repeat-containing protein are the major genes responsible for fruit color variation across the *Capsicum* species.**

Abstract Peppers (*Capsicum* spp.) rank among the most widely consumed spices globally. Fruit color, serving as a determinant for use in food colorants and cosmeceuticals and an indicator of nutritional contents, significantly influences market quality and price. Cultivated *Capsicum* species display extensive phenotypic diversity, especially in fruit coloration. Our study leveraged the genetic variance within four *Capsicum* species (*Capsicum baccatum, Capsicum chinense, Capsicum frutescens,* and *Capsicum annuum*) to elucidate the genetic mechanisms driving color variation in peppers and related Solanaceae species. We analyzed color metrics and chromatic attributes (Red, Green, Blue, *L**, *a**, *b**, Luminosity, Hue, and Chroma) on samples cultivated over six years (2015–2021). We resolved genomic regions associated with fruit color diversity through the sets of SNPs obtained from Genotyping by Sequencing (GBS) and genome-wide association study (GWAS) with a Multi-Locus Mixed Linear Model (MLMM). Significant SNPs with FDR correction were identified, within the Cytochrome P450, MYB-related genes, Pentatricopeptide repeat proteins, and ABC transporter family were the most common among the four species, indicating comparative evolution of fruit colors. We further validated the role of a pentatricopeptide repeat-containing protein (Chr01:31,205,460) and a cytochrome P450 enzyme (Chr08:45,351,919) via competitive allele-specific PCR (KASP) genotyping. Our findings advance the understanding of the genetic underpinnings of *Capsicum* fruit coloration, with developed KASP assays holding potential for applications in crop breeding and aligning with consumer preferences. This study provides a cornerstone for future research into exploiting *Capsicum*'s diverse fruit color variation.

Communicated by Sanwen Huang.

Ambika Bhattarai and Padma Nimmakayala have contributed equally to this work.

✉ Padma Nimmakayala
padma@wvstateu.edu

✉ Nagamani Balagurusamy
bnagamani@uadec.edu.mx

✉ Umesh K. Reddy
ureddy@wvstateu.edu

[1] Gus R. Douglass Institute and Department of Biology, West Virginia State University, Institute, WV, USA

[2] Laboratorio de Biorremediación, Facultad de Ciencias Biológicas, Universidad Autónoma de Coahuila, 27275 Torreon, Coahuila, Mexico

[3] Department of Biology, University of Texas Rio Grande Valley, Edinburg, TX, USA

[4] Division of Plant & Soil Sciences, West Virginia University, Morgantown, WV, USA

[5] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA

Introduction

Capsicum genus is distinguished in the culinary and nutritional world, serving as a repository of bioactive compounds. These bioactive constituents include essential vitamins like C and E, critical minerals such as potassium and magnesium, and a plethora of secondary metabolites including carotenoids like capsanthin and capsorubin, flavonoids such as quercetin and kaempferol, phenolic acids including caffeic and ferulic acids, and tannins (Azlan et al. 2022; Deepa et al. 2007; Howard et al. 2000; Mendes and Goncalves 2020). The extensive phenotypic diversity within *Capsicum* species has prompted in-depth research into the genetic bases of these compound biosyntheses (Villa-Rivera and Ochoa-Alejo 2021; Wahyuni et al. 2013), with fruit color being a particularly intriguing phenotype given its implications for ripeness, biochemical composition, and marketability, which in turn influences consumer preference and export prospects (Frank et al. 2001; Scossa et al. 2019). The modern market's penchant for unusually colored vegetables, such as peppers, has led to these varieties fetching premium prices due to consumer appeal (Schifferstein et al. 2019). Thus, targeting fruit color in breeding programs for vegetable species like chili peppers is of considerable interest.

The evolutionary history of *Capsicum* species, rooted in the Brazilian-Bolivian regions, represents a narrative of adaptation and natural selection (Walsh and Hoot 2001). The radiation of this genus across varied American ecosystems, from the lush Amazon to the arid Central America, unveils speciation intertwined with ecological diversity (Carrizo García et al. 2016; Kraft et al. 2014). In addition, the domestication journey, marked by human selection for diverse morphologies and evolutionary bottlenecks during domestication routes, has been pivotal in crafting the assortment of domesticated pepper species (Aguilar-Meléndez et al. 2009). The genetic underpinnings of traits like diverse colors evolve to favor bird-mediated seed dispersal, influencing plant survival and consumer preference (Walsh and Hoot 2001). Thus, the vibrant spectrum of *Capsicum* spp. fruit colors exemplify convergent evolution, with parallel evolutionary trajectories across diverse cultivated lineages evolving similar fruit colors to attract dispersers like birds, indicating a common evolutionary strategy for reproductive success (Pickersgill 1971).

The genetic intricacies of fruit color are not exclusive to *Capsicum* spp. but are widespread across various horticultural crops (Tong et al. 2022). The diverse mature fruit colors stem from chloroplast-to-chromoplast conversion during ripening, with chromoplasts possessing a more significant carotenoid biosynthesis and storage capacity than chloroplasts, leading to a spectrum of colors in mature pepper fruits (Kapoor et al. 2022b). Notable genes like phytoene synthase 1 (*PSY1*) and capsanthin-capsorubin synthase (*CCS*) are known to be pivotal in carotenoid biosynthesis, influencing pepper fruit color variation (Borovsky and Paran 2008; Chen et al. 2023; Hurtado-Hernandez and Smith 1985; Sim et al. 2012). However, the role of anthocyanins, chlorophyll, flavonoids, and the regulatory genes involved in these pathways remains unclear (Nabi et al. 2023). For example, research on tomatoes has shed light on the *SlMYB12* gene's influence on flavonoid biosynthesis and fruit coloration (Ballester et al. 2016), while grape studies have demonstrated the significance of *VvMYBA1* and *VvMYBA2* genes in controlling anthocyanin accumulation and fruit skin color (Fournier-Level et al. 2009). Moreover, in apples, the *MdMYB10* gene has been linked to anthocyanin levels and the red hue of the fruit (Chagne et al. 2013). These findings underscore the complex genetic network dictating fruit color, highlighting the necessity for more extensive research in *Capsicum* spp. (Wang et al. 2023).

While several GWAS have been conducted on *Capsicum* collections, most have focused on various fruit traits of importance (McLeod et al. 2023; Nimmakayala et al. 2021). Our study leverages a broad and varied collection of cultivated *Capsicum* species to investigate the genetic underpinnings of fruit color diversity. It is essential to recognize that fruit color is a complex, polygenic trait, likely governed by a network of genes and transcription factors (Rodriguez-Uribe et al. 2012; Wang et al. 2023). Thus, the current research focuses on understanding the evolution of shared fruit color components in cultivated *Capsicum* species.

Materials and methods

Plant material and growth condition

Capsicum collections were obtained from our historical collections, USDA-ARS, Germplasm Resource Information Network, Plant Genetic Resources Conservation Unit, Griffin, GA, and World Vegetable Center (AVRDC, Shanhua, Taiwan). Our diverse collection encompasses 665 accessions that cut across four cultivated *Capsicum* species. Specifically, the dataset includes 159 accessions of *Capsicum annuum*, 196 of *Capsicum baccatum*, 167 of *Capsicum chinense*, and 143 of *Capsicum frutescens*; origin details for each accession are provided in Supplementary Table S1. For *C. annuum, C. baccatum,* and *C. chinense*, five plants per accession were used from greenhouse and field conditions across three consecutive years: 2013, 2014, and 2015. Meanwhile, the *Capsicum frutescens* collection was separately grown in 2020, adhering to standard horticultural production practices.

Genomic DNA extraction, GBS, and SNPs calling

High-quality genomic DNA was isolated from these samples using the DNeasy Plant Mini Kit (QIAGEN, Germany), following the manufacturer's guidelines. Following DNA isolation, the samples were treated with *Ape*KI, a Type II restriction endonuclease, to digest the genomic DNA. The fragmented DNA was ligated to barcode-specific adapter pairs post-digestion to facilitate subsequent sequencing steps. The prepared DNA libraries were then sequenced on an Illumina Nextseq500 platform (Illumina Inc., USA), employing a methodology adapted from (Elshire et al. 2011). This high-throughput sequencing approach generated Genotyping-By-Sequencing (GBS) reads, which were meticulously aligned to the reference *Capsicum* genomes (*Capsicum annuum* v. 1.55; *Capsicum baccatum* v. 1.2; *Capsicum chinense* v. 1.2) available at the Pepper Genome Database (http://pgd.pepper.snu.ac.kr/index.php?a=view) using BWA-MEM, as described in (Kim et al. 2017, 2014). We used the GB-easy tool to identify Single Nucleotide Polymorphisms (SNPs) (https://github.com/dpwickland/GB-eaSy). This tool facilitated the automated calling of SNPs from the aligned GBS reads, executed through a pre-established command pipeline. The final output of this analytical workflow was a Variant Call File (VCF), primed for downstream analyses and further investigations. A SNP density plot was generated by using CMplot (https://www.bioinformatics.com.cn/plot_basic_SNP_density_by_CMplot) to observe the SNP distribution among chromosomes for each specie.

Phenotyping

Ripened pepper fruits, reaching their peak maturity approximately 60 days post-flowering, were carefully harvested for color analysis. Digital images of the whole, uncut fruits were captured with meticulous precision using an Epson Workforce GT-1500 Color Document Scanner, set to a high-resolution scanning mode to ensure accurate color capture. These high-fidelity JPEG images were then imported into the specialized Tomato Analyzer software, which provides an accurate, high-throughput way of quantifying diverse fruit morphological attributes, including color (Darrigues et al. 2008). This software was used in prior studies to measure variation in color attributes in fruits of tomato and pepper (Nankar et al. 2020a, 2020b; Robarts et al. 2012). Within the software, the color characteristics of the *Capsicum* spp. fruits were analyzed and quantified across multiple color representation systems, including RGB (Red–Green–Blue), CIELAB (International Commission on Illumination color space), and the Munshell color system (Munshell 1912). This comprehensive analysis generated extensive numerical data points, including RGB values, luminosity, and coordinates in the CIELAB color space (L^*, a^*, b^*) and Hue and Chroma values (that is, 9 color dimensions, see Table S2 for detailed definitions of fruit color measurements). These metrics provided an exhaustive and accurate representation of the fruit's color characteristics. This dataset is a critical resource for subsequent GWAS, providing an empirical foundation for understanding the genetic underpinnings of fruit color variation in *Capsicum* species.

GWAS

In this study, we considered only SNPs successfully mapped to known genomic locations within the reference genomes to eliminate the possibility of spurious associations. These mapped SNPs underwent additional quality controls, including a Minor Allele Frequency (MAF) threshold of 0.05 (Weale 2010) and a call rate of 90% to create a high-confidence dataset. To account for the confounding effects of population stratification, we incorporated population structure and kinship matrices as covariates in our analyses. Principal Component Analysis (PCA) was conducted using the SNP & Variation Suite (SVS v8.9.1) software Golden Helix, Inc. The first three principal components explained most of the observed genetic variation and were integrated into the GWAS. Marker-trait associations were screened through a Multi-Locus Mixed Linear Model (MLMM), optimizing the reliability and robustness of the detected associations. The significance of these associations was visually illustrated through Quantile–Quantile (Q-Q) and Manhattan plots. We established a significance threshold based on false discovery rate (FDR) correction. The chromosomal distribution of the associated SNPs identified in the GWAS was mapped using Circos software (http://circos.ca/) with the SNP region magnified to 2 Mb across *Capsicum* genomes.

Candidate gene selection and functional annotation

Gene annotation is pivotal in deciphering the molecular underpinnings of fruit color variations. We employed a comprehensive search through gene annotation databases to identify candidate genes implicated in these variations. We utilized SNPEff, an advanced bioinformatics tool, to rigorously analyze the SNPs associated with fruit color variations. This tool locates the SNPs within the genome. It predicts the functional impact of these variants based on the coding effects they exert on the reference genome, providing a nuanced understanding

of their role in color variation (Cingolani 2022). Our analysis encompassed various genomic regions, including genic and intergenic regions, exons, introns, and splice junctions. Candidate genes from these annotated regions were examined across *Capsicum annuum v. 1.55; Capsicum baccatum v. 1.2; Capsicum chinense v. 1.2* genomes. This cross-species comparison facilitated the identification of a core set of genes universally implicated in determining fruit color, thereby offering invaluable insights into the conserved genetic mechanisms governing this complex trait.

KEEG pathway and gene expression profile of significantly associated genes

Pathway mapping of significantly associate genes was carried out using KOBAS (http://kobas.cbi.pku.edu.cn/) and the KEGG database. Enriched KEGG pathways were displayed by using the GOseq R package. RNA-seq gene expression data for significantly associated genes in *C. chinense* were retrieved from our previously published work (Natarajan et al. 2020). The Fragments Per Kilobase transcripts per Million mapped reads (FPKM) values for three *C. chinense* accessions Naga morich, PI 224448, and PI 257129, showing different fruit color, were used to generate a heatmap and compare the expression of associated genes identified.

Kompetitive allele-specific PCR (KASP)

Sequences flanking selected SNPs were analyzed by the KASP™ assay design tool of Biosearch Technologies in California, USA. The KASP assay was developed per the manufacturer's protocol (He et al. 2014). Each reaction was prepared with 5.0 µl of KASP Master Mix—comprising two allele-specific forward primers, one reverse primer, universal fluorescent probes, Taq polymerase, and dNTPs in an optimized buffer solution—and 0.138 µl of the Assay Mix. To this mixture, 5 µl of DNA was added. The polymerase chain reaction (PCR) was performed in three stages: an initial cycle at 94 °C for 15 min, followed by ten cycles of denaturation at 94 °C for 20 s, and annealing/extension from 61 to 55 °C decreasing by 0.6 °C per cycle for 60 s each, and concluding with 35 cycles at 94 °C for 20 s and a constant 55 °C for 60 s for annealing/extension. The genotyping results were analyzed using the StepOne plus auto-caller software from Applied Biosystems.

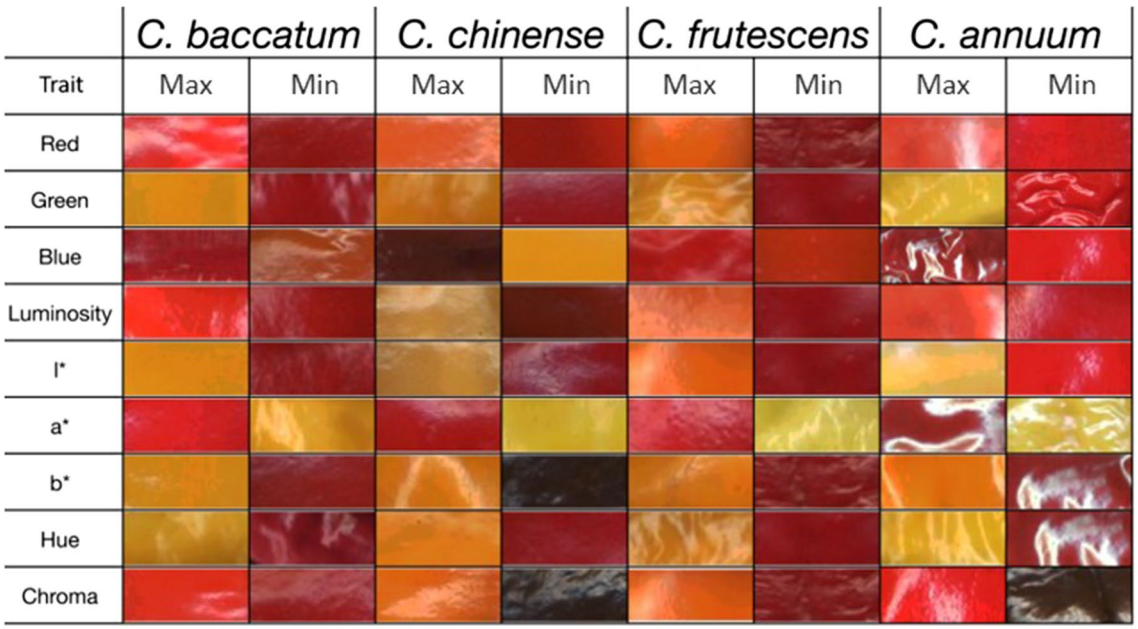
Results

Genotyping analysis of *Capsicum* species

In total, 220, 226, 188, and 162 accessions for *C. baccatum*, *C. annuum*, *C. chinense*, and *C. frutescens* were genotyped using genotyping by sequencing. A range from 128.1 to 698.6 million reads was generated for *C. annuum* and *C. chinense*, respectively. The percentage of mapped reads ranged from 84 to 93% for all four species to their respective genome (Table S3). Further, utilizing the GBeasy pipeline, we generated a robust set of SNPs for each *Capsicum* species under study. Specifically, we identified 698,636 SNPs in *C. annuum*, 2,143,813 in *C. baccatum*, 301,884 SNPs in *C. chinense*, and 2,335,414 SNPs in *C. frutescens*. To ensure the reliability and quality of our dataset, we applied stringent filtering criteria, including a minor allele frequency (MAF) threshold of 0.05 and a call rate greater than 90%. After using these filters, the resulting high-confidence SNP sets were narrowed down to 11,098 for *C. annuum*, 26,852 for *C. baccatum*, 42,410 for *C. chinense*, and 57,741 for *C. frutescens*. This refinement in SNP numbers ensures that the subsequent analyses are based on a high-quality, reliable dataset, thereby bolstering the robustness of our findings. Overall, SNPs were distributed across all chromosomes in all *Capsicum* species, with *C. annuum* showing the lowest number of polymorphisms in chromosome 08 (Figure S1). The distribution of SNPs by chromosome is in Table S3.

Phenotypic variation

In our study, we examined a diverse *Capsicum* population encompassing 665 accessions from four cultivated species of *Capsicum*. Our objective was to gain an in-depth understanding of the phenotypic variation in fruit color across these species. We employed the Tomato Analyzer software to quantitatively assess three color spaces, leading to nine color dimensions—Red, Green, Blue, L^*, a^*, b^*, Luminosity, Hue, and Chroma (phenotypic data for each accession among all *Capsicum* species is provided in the Supplementary Tables S4-S7). Our analysis revealed a rich tapestry of phenotypic diversity across these color metrics. Continuous and extensive variations were observed for each of the nine color dimensions within our *Capsicum* accessions, representing all four species under study (Fig. 1 and Figure S2). This diversity is illustrated in box plots presented in Fig. 2. In addition to the box plots, the frequency distribution of these nine color dimensions was analyzed and depicted in Figure S3. The data exhibit a



	C. baccatum		*C. chinense*		*C. frutescens*		*C. annuum*	
Trait	Max	Min	Max	Min	Max	Min	Max	Min
Red	215.86	93.92	211.54	72.27	211.62	124.38	232.20	121.00
Green	145.55	31.82	147.33	30.26	197.48	34.29	189.40	37.60
Blue	73.22	26.60	116.74	27.84	180.89	27.17	139.90	35.50
Luminosity	131.65	66.57	127.70	55.43	183.90	76.14	160.20	79.20
L*	59.77	25.09	60.38	23.74	74.81	27.42	72.70	29.50
a*	54.92	11.13	51.06	6.18	51.53	5.48	59.40	8.80
b*	67.02	12.13	61.33	4.28	63.51	20.68	70.30	4.10
Hue	186.74	19.03	87.40	19.71	205.40	27.32	342.70	15.60
Chroma	71.04	29.14	66.47	10.15	74.17	17.97	75.30	15.10

Fig. 1 Manifestation of individual color dimensions derived using CIELAB and Munsell color system in various *Capsicum* species indicating the accession name with maximum and minimum color values for each species

normal distribution pattern, suggesting a complex genetic architecture underlying these traits across the four *Capsicum* species. This evidence further underscores the importance of our research in unraveling the genetic intricacies of fruit color variation within the *Capsicum* genus.

Correlation network analyses

Our study used correlation network analysis to determine the relationships among various color traits across multiple *Capsicum* species. Utilizing the 'corrplot' package in the statistical software R, we constructed intricate correlation networks, the results of which are graphically represented in Figure S4. The analysis revealed intriguing patterns of associations between the nine color dimensions in the three color spaces, some of which were species-specific. For instance, within the *C. baccatum* accessions, nearly all color dimensions displayed positive correlations except for color dimension *a**. This metric showed a pronounced negative correlation with green and hue while maintaining weak associations with other color dimensions. In contrast, the *C. chinense* accessions revealed that color dimension *a** had strong negative correlations with green, hue, and *L**. Moreover, the color dimension blue in *C. chinense* was negatively correlated with both chroma and luminosity, emphasizing the species-specific variations in color associations. *C. frutescens* collection also exhibited compelling patterns, with *a** displaying strong negative associations with green, blue, hue, and chroma. This color dimension showed negligible to weak correlations with the remaining color dimensions. In the *C. annuum* collections, the color dimension *a** was negatively correlated with green, *L**, blue, and hue. Significantly, *a** trait positively correlated with chroma in this species. Furthermore, color dimension *b** in *C. annuum* presented an intriguing divergence, showing little to no correlation

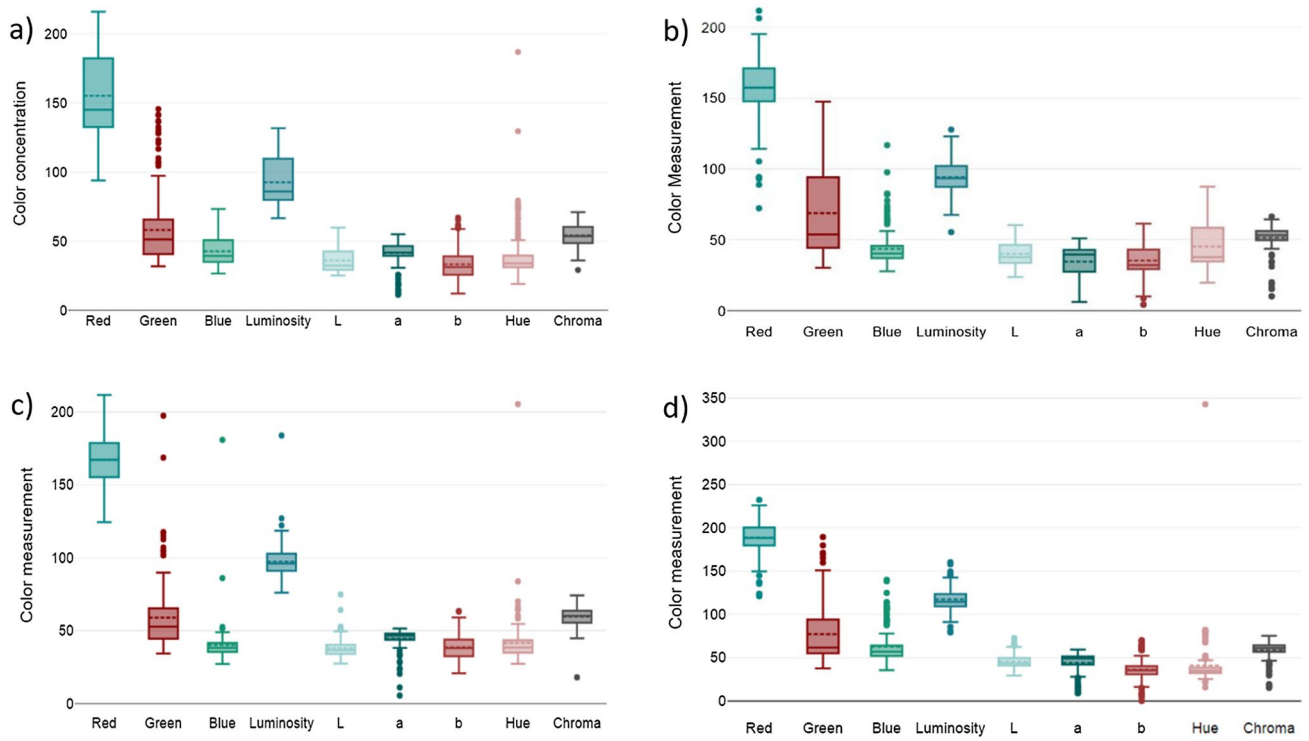


Fig. 2 Range for various color dimensions among the accessions belongs to **a** *C. baccatum*, **b** *C. chinense,* **c** *C. frutescens,* and **d** *C. annuum*

with the other color dimensions. These intricate patterns of correlation provide valuable insights into the multifaceted genetic architecture that governs fruit color across different *Capsicum* species, and they underscore the complexity of these traits.

Population stratification

We analyzed to comprehensively dissect the genetic architecture underlying our extensive collection of 665 *Capsicum* accessions. Initially, we employed principal component analysis (PCA) on a rigorously quality-controlled SNPs dataset. This analytical approach was a powerful tool for evaluating genetic diversity and capturing the nuances of population stratification within our *Capsicum* cohorts. As depicted in Figure S5, the genomic PCA plots for all *Capsicum* species exhibited a diffuse scattering of accessions, underscoring the existence of intricate population substructures based on fruit colors. Building upon the PCA, we subsequently conducted Numeric Principal Analyses across all color spaces under investigation. This approach aimed to scrutinize the efficacy of these color spaces in delineating the distinct fruit color variations among the accessions. The numeric PCA plots manifested demarcated clusters, each corresponding to specific fruit colors (Fig. 3). As illustrated in numeric PCA, these results unambiguously demonstrate that the chosen color dimensions serve as robust indicators, effectively capturing the intrinsic color variations among different *Capsicum* species.

GWAS and candidate genes identification for fruit color trait

To unravel the complex genetic architecture driving fruit color diversity across *Capsicum* species, we employed a comprehensive association mapping strategy utilizing three distinct genetic models—additive, dominant, and recessive. These models were applied across various color dimensions: Red, Green, Blue, Luminosity, *L**, *a**, *b**, Hue, and Chroma. Our rigorous statistical framework identified many significantly associated loci, characterized by -log10 P-value exceeding 3. The resulting SNP-trait associations were visualized through Manhattan plots, tailored for each color dimension (See Fig. 4 for one example using Luminosity). Manhattan plots for selected SNP markers for color traits are presented in supplementary Figures S6-S13 for the other eight color dimensions. Specifically, we discovered 263 loci in *C. annuum*, 545 in *C. baccatum*, 623 in *C. chinense*, and 465 in *C. frutescens*. To refine this expansive list and zero in on the most consequential genetic markers, we implemented a False Discovery Rate (FDR) correction. This stringent criterion reduced the list to 40 pivotal associations for *C. annuum*, 35 for C. *baccatum*, 61 for

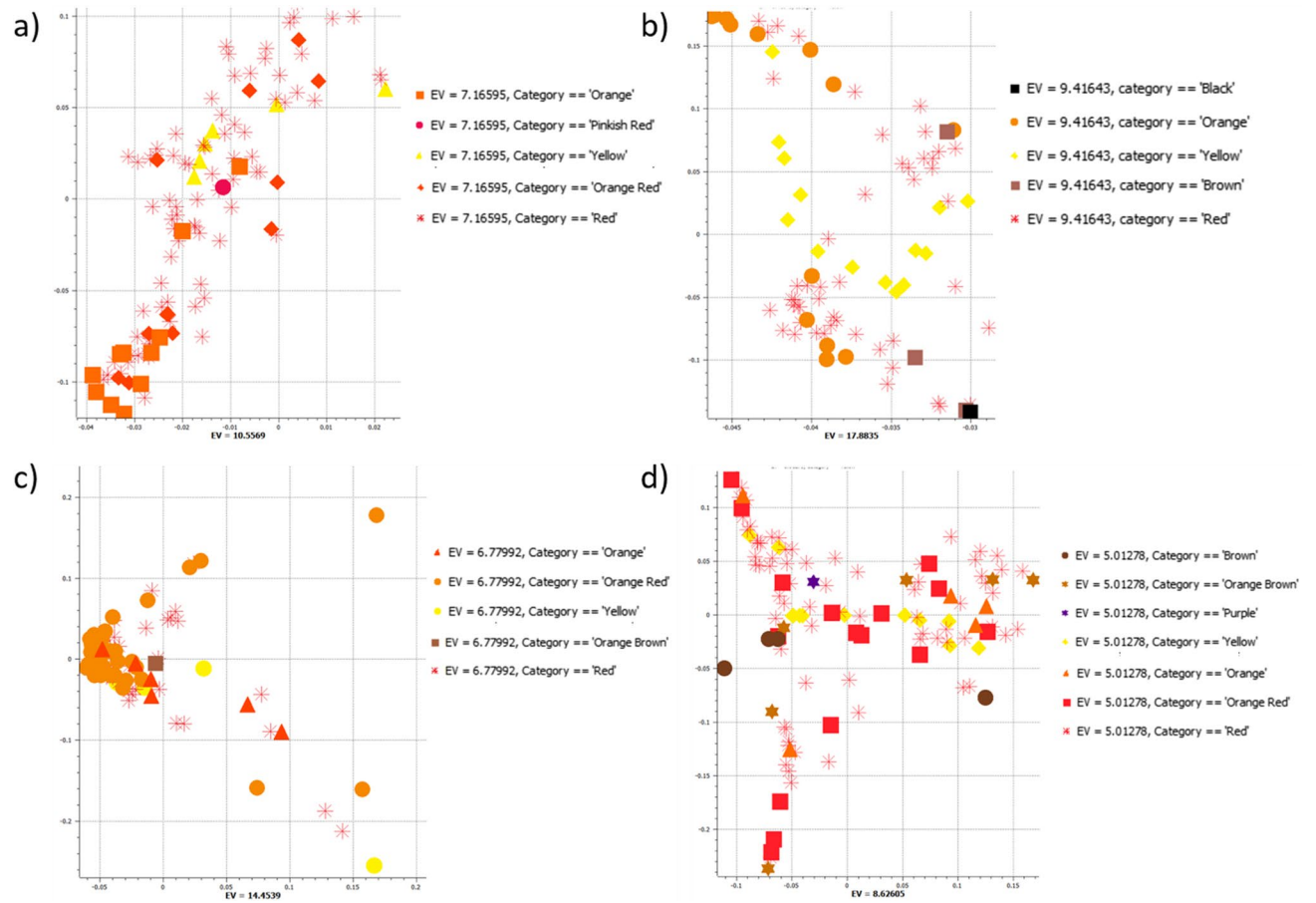


Fig. 3 Principal component analysis made using numerical values of color dimensions showing the distribution of fruit colors among the accessions belonging **a** *C. baccatum*, **b** *C. chinense,* **c** *C. frutescens,* and **d** *C. annuum*

C. chinense, and 42 for *C. frutescens*. In addition, a circus plot showing a comprehensive genomic overview of the associated SNPs distribution across three *Capsicum* species: *Capsicum annuum*, *Capsicum chinense*, and *Capsicum baccatum* is presented in Fig. 5. Association statistics, including metrics like the *P*-value, -log10 *P*-value, predictor beta, FDR correction, proportion of variability explained, Manhattan category, and primary and minor allele frequencies, is presented in Table S8. After this identification phase, we delved into functional annotations to predict candidate genes implicated in fruit color variance. Utilizing SNPeff for in-depth analyses, we pinpointed an array of likely causative SNPs and candidate genes that could be instrumental in modulating color variations. Detailed annotations of these high-impact SNPs are presented in Table S9.

Red

The comprehensive genetic analysis of four *Capsicum* species—*C. baccatum, C. chinense, C. frutescens,* and *C. annuum*—revealed a variety of significant SNPs associated with the trait Red. In *C. baccatum*, six key SNPs were identified after stringent False Discovery Rate (FDR) correction, with one SNP on chromosome 8 accounting for an impressive 23% of observed phenotypic variance. Similarly, seven significant SNPs were found in *C. chinense*, explaining a broad range of phenotypic variance from 11% to a remarkable 40%; notably, one SNP on chromosome 7 linked to a pentatricopeptide repeat-containing protein accounted for 40% of the observed variance. In the case of *C. frutescens*, four pivotal SNPs withstood FDR correction and were mapped to four different chromosomes; the range of phenotypic variance explained by these SNPs was notably wide, from 9 to 52%, with an SNP at position 11:24,776,180 being particularly impactful, explaining 52% of the observed variance. For *C. annuum*, five SNPs passed the FDR correction and were distributed across chromosomes 1, 3, and 5, explaining a phenotypic variance range of 8–25%; notably, the SNP at position 03:251,878,883, linked to a MYB-related protein, accounted for 25% of the observed variance.

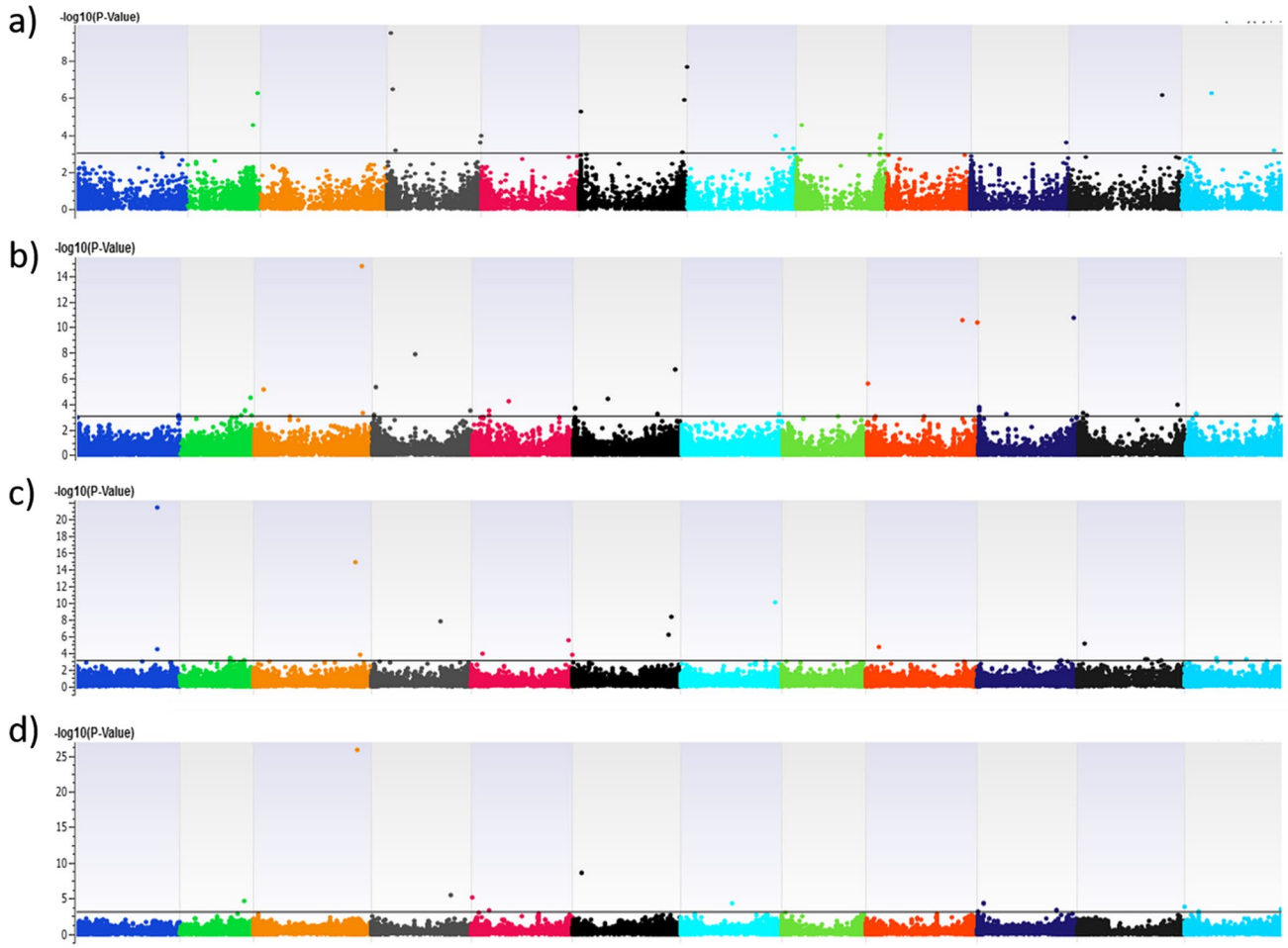


Fig. 4 Manhattan plot showing SNPs significantly associated with luminosity trait in **a** *C. baccatum*, **b** *C. chinense,* **c** *C. frutescens,* and **d** *C. annuum*. Different colors on the x-axis indicate 12 chromosomes of the pepper genome. SNPs above the horizontal line are after false discovery rate (FDR) correction. Values on *Y* axis are *P* values of –log 10

Green

In the case of *C. baccatum*, five impactful SNPs were pinpointed across chromosomes 5, 6, 8, and 12, which were responsible for explaining phenotypic disparities ranging from 7–18%. One SNP located at the 228,959,560 position on chromosome 12, in particular, was associated with a putative MYB family protein. For *C. chinense*, six prominent SNPs were found on chromosomes 1, 2, 4, 5, and 7, explaining between 11 and 24% of phenotypic variance. SNPs on chromosomes 2 and 5 were markedly linked with genes for serine/threonine-protein kinase and ABC transporter A family member 12-like, respectively. In *C. frutescens*, eight noteworthy SNPs were distributed among chromosomes 3, 5, 6, 10, 11, and 12, accounting for 9–13% phenotypic variance. These SNPs were recurrently linked to pentatricopeptide repeat-containing proteins. Finally, in *C. annuum*, five critical SNPs that passed FDR correction were identified on chromosomes 1, 6, 11, and 12, accounting for a variance from 8 and 13% in phenotype. An SNP at the 3,787,311 positions on chromosome 6 was especially significant and was associated with a pentatricopeptide repeat-containing protein gene.

Blue

In *C. baccatum*, six impactful SNPs distributed across chromosomes 1, 6, and 12 were responsible for phenotypic variance between 6 and 19%. For instance, the SNP located at position 01:31,205,460 on chromosome 1, related to a pentatricopeptide repeat-containing protein, explained an 8% phenotypic variance. In *C. chinense*, eight noteworthy SNPs spread across chromosomes 1, 3, 4, 5, 6, 10, and 11 accounted for 11 to 17% variance. SNPs on chromosome positions 03:6,730,447 and 10:28,733,075 were significantly associated with F-box/kelch-repeat proteins. In the case of *C. frutescens*, seven

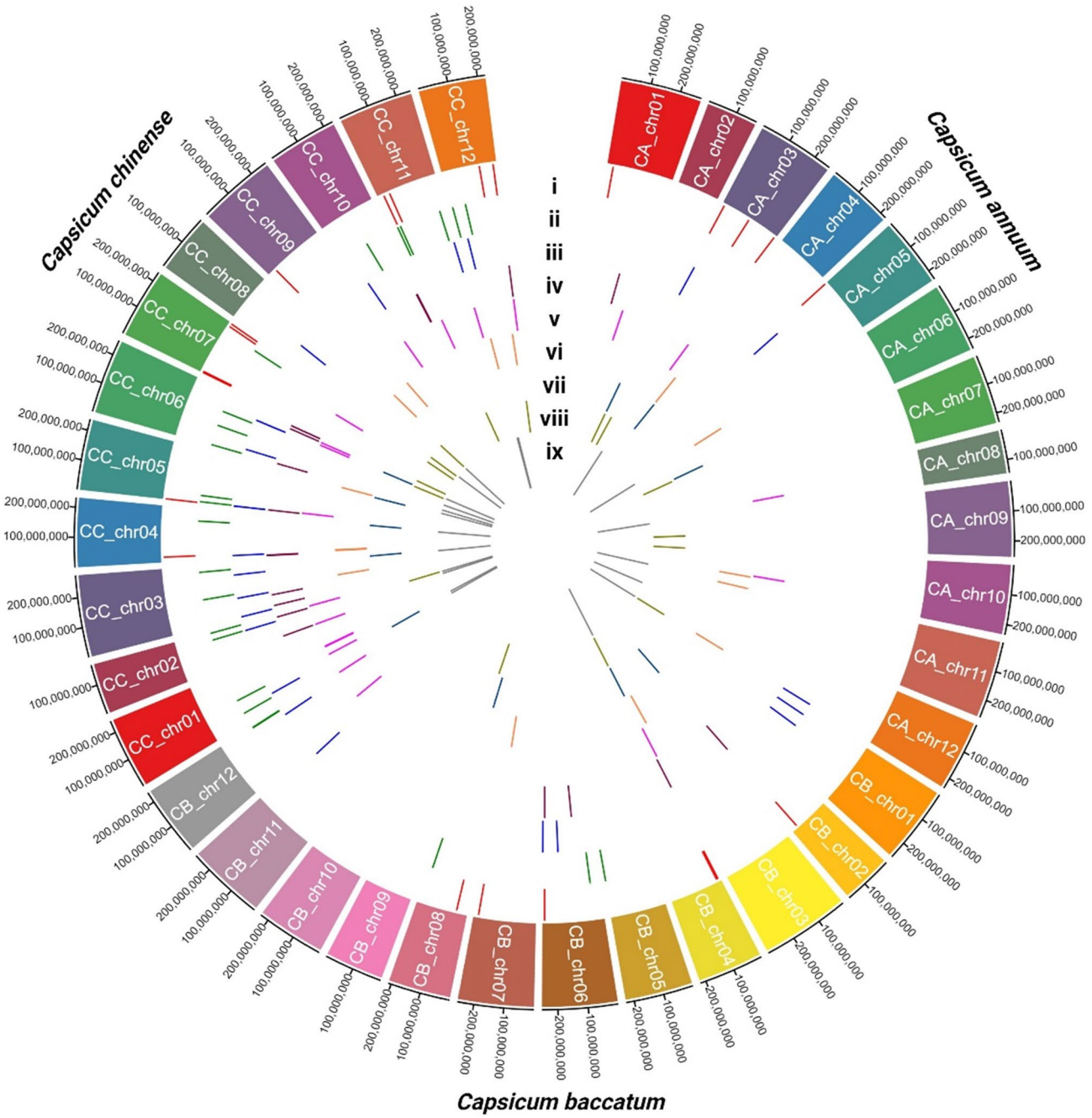


Fig. 5 Collinearity of SNP positions associated with various color traits in *Capsicum* species. Nine concentric layers of SNPs indicate how a given SNP is associated to multiple color dimensions. Each of the nine layers are placed in the order of **i** red, **ii** green, **iii** blue, **iv** luminosity, **v** L, **vi** a, **vii** b, **viii** hue, and **ix** chroma (color figure online)

pivotal SNPs were located on chromosomes 2, 3, 6, 8, and 12, and they accounted for phenotypic variance from 9 to 13%. These SNPs were recurrently linked to pentatricopeptide repeat-containing proteins, especially in the blue color dimension. Finally, for *C. annuum*, three SNPs that passed the FDR correction were located on chromosomes 3, 5, and 12, explaining the phenotypic variance of 10 to 14%. Specifically, the SNP at position 03:24,306,417 on chromosome 3 was associated with a protein FAR1-RELATED SEQUENCE 11-like gene and was particularly noteworthy for blue color variance.

Luminosity

For *C. baccatum*, four standout SNPs across chromosomes 2, 4, and 6 explained phenotypic variance between 8 and 21% (Fig. 4). Notably, the SNP on chromosome 04:8,593,005 was linked to a calpain-type cysteine protease, whereas the SNP at 02:148,766,979 was associated with cyclin D3.1. Turning to *C. chinense*, eight prominent SNPs spread across chromosomes 2, 3, 4, 5, and 10 accounted for a variance from 11% to a significant 33%. Remarkably, the SNP on chromosome 03:84,522,766 was linked with an arginine/serine-rich coiled-coil protein 2 isoform X2 and a MYB-like protein I. Additionally, SNPs on chromosomes 04:7,751,133, 05:3,142,405, and 10:216,624,158 were all tied to probable LRR receptor-like serine/threonine-protein kinases, and in the case of *C. frutescens*, six significant SNPs spanned chromosomes 2, 6, and 12. Pentatricopeptide repeat-containing proteins were especially prominent, with SNPs on chromosome positions 02:41,407,693 and 06:3,822,587 explaining 11 and 12% phenotypic variance, respectively. Notably, SNPs at chromosome 06:230,103,252 and 06:230,103,244 were related to ABC transporter G family members, which shed light on their genetic significance. Lastly, in *C. annuum*, a singular standout SNP at chromosome position 01:169,031,114, associated with a zinc finger protein 598-like gene, explained 8% of the observed phenotypic variance.

*L**

For *C. baccatum*, three key SNPs located on chromosomes 4 and 12 accounted for 13 and 6% of phenotypic variance, respectively. Specifically, an SNP at chromosome position 12:39,093,634 was related to a calcium-dependent protein kinase, while another SNP at 12:228,886,503 was linked to an ABC transporter B family member. An additional SNP at 04:128,593 was associated with a probable LRR receptor-like serine/threonine-protein kinase. In *C. chinense*, eight significant SNPs were spread across chromosomes 1, 2, 5, 6, 10, 11, and 12, influencing phenotypic variance ranging from 12% to an astonishing 76%. Notable SNPs were located at chromosome positions 01:153,341,594 and 01:98,936,335, linked to cinnamoyl-CoA reductase and serine/threonine-protein phosphatase 4 regulatory subunit. Additional SNPs were associated with zinc finger domain-containing stress-associated proteins and F-box/kelch-repeat proteins. For *C. frutescens*, five SNPs that withstood FDR correction were dispersed across chromosomes 1, 2, 6, 7, and 11, explaining a 10 and 14% variance. Noteworthy were SNPs related to pentatricopeptide repeat-containing proteins at various chromosome positions and an SNP at 06:210,658,508 linked to a probable LRR receptor-like serine/threonine-protein kinase. Finally, in *C. annuum*, four SNPs passed FDR correction located on chromosomes 1, 3, 8, and 10. These explained phenotypic variance ranging from 8–10%. The SNP on chromosome 01:252,255,946 was particularly associated with a pentatricopeptide repeat-containing protein. At the same time, other significant SNPs were linked to gibberellin 2-beta-dioxygenase, 26S proteasome non-ATPase regulatory subunit, and a nuclear transcription factor.

*a**

In *C. baccatum*, only two SNPs, located on chromosomes 3 and 7, withstood the FDR correction among 52 associated SNPs. These SNPs accounted for phenotypic variance of 8% and 12%. Specifically, the SNP at position 03:279,762,481 was linked to a probable LRR receptor-like serine/threonine-protein kinase, while the one at 07:210,731,968 was in an intergenic region associated with isocitrate lyase and MYB-like transcription factor ETC3. For *C. chinense*, five SNPs emerged as significant across chromosomes 4, 5, 9, and 12, explaining phenotypic variance ranging from 11 to 12%. For example, the SNP at position 09:133,866,441 was associated with a probable F-box protein and ubiquinol oxidase 2, while others were linked to various protein kinases and ABC transporters. In *C. frutescens*, three significant SNPs were found on chromosomes 3, 9, and 12, all accounting for a phenotypic variance of 10%. Notably, SNPs on chromosomes 12 and 9 were linked to pentatricopeptide repeat-containing proteins, while the one on chromosome 3 at position 03:135,620,678 was associated with peroxidase 5-like, explicitly affecting the color dimension *a**. Lastly, for *C. annuum*, six SNPs stood out among 46, located on chromosomes 1, 4, 6, 10, 11, and 12. These SNPs explained substantial phenotypic variance, ranging from 18–31%. Several of these SNPs were linked to various types of protein kinases, heat shock proteins, and other functional proteins, indicating their significant roles in phenotypic expression.

*b**

In *C. baccatum*, of the 89 identified SNPs, three on chromosomes 2, 4, and 8 were significantly associated with phenotypic variance ranging from 9–15%. Specifically, the SNP at chromosome 08:144,332,220 was linked to phospholipase D Z-like and histidine decarboxylase-like genes. In contrast, others were associated with putative auxin efflux carriers and calpain-type cysteine proteases. For *C. chinense*, four SNPs across chromosomes 3, 4, 5,

and 6 emerged as significant from 34 identified, explaining an extensive range of phenotypic variance from 10 to 54%. For example, the SNP at chromosome 05:217,136,990 was a cofactor for b* variance associated with ras-related and BTB/POZ domain-containing proteins. Another SNP at chromosome 04:224,586,097, linked to an F-box protein, was also noteworthy. In *C. frutescens*, out of 52 associated SNPs, only one on chromosome 1 was significant, explaining a 10% phenotypic variance. This SNP, located at 01:53,151,302, was in the intergenic region between E3 ubiquitin-protein ligase UPL6-like and pentatricopeptide repeat-containing protein genes. Lastly, in *C. annuum*, three significant SNPs were identified among 51, located on chromosomes 2, 4, and 6, accounting for phenotypic variance between 9 and 37%. For instance, the SNP at chromosome 04:13,533,319 was associated with a putative receptor-like protein kinase, while others were linked to subtilisin-like protease and NAC domain-containing protein genes.

Hue

In *C. baccatum*, among 156 identified SNPs, two were significant, located on chromosomes 4 and 8, accounting for phenotypic variance of 20 and 16%, respectively. Specifically, the SNP on chromosome 04:15,303,598 was in an intergenic region linked to (RS)-norcoclaurine 6-O-methyltransferase-like and probable LRR receptor-like serine/threonine-protein kinase. Another significant SNP at 08:193,074,112 was a cofactor for variance in Pentatricopeptide repeat-containing proteins. For *C. chinense*, six SNPs emerged as significant from a pool of 163, located on chromosomes 2, 6, 7, 8, 11, and 12. These explained phenotypic variance ranging from 11% to an extraordinary 80%. Noteworthy SNPs were located at chromosome 07:211,354,155, associated with an F-box protein, and 12:224,304,164, linked to a probable serine/threonine-protein kinase. In *C. frutescens*, three significant SNPs were found among 277 associated, located on chromosomes 6 and 8, accounting for phenotypic variance between 11 and 15%. SNP at 06:2634 was associated with an ABC transporter B family member, whereas others were linked to Pentatricopeptide repeat-containing proteins. Lastly, in *C. annuum*, seven SNPs were identified as significant among 75, located on chromosomes 2, 3, 6, 9, 10, and 12, and contributing to phenotypic variance ranging from 10–44%. For instance, the SNP at 12:234,434,115 was associated with a calcium-dependent protein kinase. In contrast, others were linked to various proteins, including GLABROUS1 enhancer-binding proteins and F-box/kelch-repeat proteins, and were found to be cofactors for the Hue color dimension.

Chroma

For *C. baccatum*, out of 44 associated SNPs, four remained significant after FDR correction. These were located on chromosomes 4, 8, 11, and 12, accounting for 7 and 31% of phenotypic variance, respectively. Key SNPs were found at positions such as 04:10,832,611, linked to a probable LRR receptor-like serine/threonine-protein kinase, and 08:45,351,919, associated with cytochrome P450 CYP72A219-like. In *C. chinense*, nine significant SNPs were discovered among 89 associated SNPs on chromosomes 1, 2, 4, 6, and 12. These explained phenotypic variance from 10 to 21%. Notable SNPs included those at chromosome positions 02:132,385,192 and 01:53,651,476, both associated with F-box proteins. For *C. frutescens*, five SNPs remained significant from a set of 52 associated SNPs. These were located on chromosomes 3, 6, 8, and 12, accounting for a notably wide range of phenotypic variance, from 10 to 62%. For example, an SNP at 03:243,301,602 was associated with a nuclear transcription factor, while another at 06:19,381,223 was linked to actin-depolymerizing factor 2-like. Finally, in *C. annuum*, six significant SNPs were identified among 35 associated SNPs distributed across chromosomes 3, 6, 8, 11, and 12. These accounted for phenotypic variance ranging from 8–26%. For instance, two SNPs at positions 12:232,533,746 and 12:1,969,418 were associated with F-box/kelch-repeat proteins and accounted for a phenotypic variance of 8%.

Comparative analysis of GWAS results

Furthermore, based on the annotation of genes containing associated SNPs related to various color traits, we identified several common candidate genes with similar functions associated with these traits. A complete list of SNP markers showing pleiotropic effect among color traits is listed in Table 1 and Fig. 5. Notably, the SNPs generally explain a range of phenotypic variance across all species and traits, from as low as 6% to as high as 80%. For instance, 17 genes encoding pentatricopeptide repeat-containing protein were observed in all color traits except for the *b** trait; these proteins were also associated with all *Capsicum* species. In addition, we found nine genes with an LRR receptor-like serine/threonine-protein kinase function and were more commonly associated with *C. chinense, C. baccatum*, and *C. frutescens,* while they are less prevalent in *C. annuum*. Moreover, F-box/kelch-repeat protein function was found on seven genes related to all *Capsicum* species except *C. baccatum*. Genes with

Table 1 Common gene annotations shared between fruit color traits among *Capsicum* species

Color trait	SNP position	Gene ID	Annotation
Green	01:42,272,375	CC.CCv1.2.scaffold549.1	ABC transporter A family
Chroma	12:232,533,746	CA12g20900	ABC transporter B family
Hue	06:2634	CC.CCv1.2.scaffold1019.1	
*L**	12:228,886,503	CB.CBv1.2.scaffold2474.1	
Blue	02:132,022,079	CC.CCv1.2.scaffold823.48	
Chroma	08:26,089,056	CC.CCv1.2.scaffold183.46	
*a**	04:20,870,037	CC.CCv1.2.scaffold302.16	ABC transporter G family
Luminosity	06:230,103,252	CC.CCv1.2.scaffold31.72	
Red	05:1,435,954	CC.CCv1.2.scaffold856.8	ABC transporter I family
Blue	12:38,149,033	CC.CCv1.2.scaffold201.30	Callose synthase 5
Chroma	12:38,149,046	CC.CCv1.2.scaffold201.30	
*b**	04:8,335,730	CB.CBv1.2.scaffold1012.15	Calpain-type cysteine protease DEK1
Luminosity	04:8,593,005	CB.CBv1.2.scaffold1012.15	
Green	12:6,341,820	CA12g02930	Cytochrome P450
Red	12:155,572,462	CC.CCv1.2.scaffold58.22	
	08:45,351,919	CB.CBv1.2.scaffold473.5	
Blue	03:6,730,447	CC.CCv1.2.scaffold647.51	F-box/kelch-repeat protein At3g23880-like
Blue	03:6,730,447	CC.CCv1.2.scaffold647.52	
Blue	10:28,733,075	CC.CCv1.2.scaffold97.1	
Chroma	12:1,969,418	CA12g00600	
Hue	03:17,947,103	CA03g06280	
*L**	10:10,391,356	CC.CCv1.2.scaffold1234.4	
*L**	12:226,262,158	CC.CCv1.2.scaffold710.69	
Luminosity	12:224,920,669	CC.CCv1.2.scaffold1389.1	
b*	06:192,909,983	CC.CCv1.2.scaffold501.1	Flavin-containing monooxygenase FMO GS-OX-like 4
Chroma	06:192,759,203	CC.CCv1.2.scaffold501.5	
*b**	06:205,140,356	CA06g14960	NAC domain-containing protein 78-like
Chroma	12:33,932,768	CB.CBv1.2.scaffold2291.2	
Chroma	05:233,972,444	CC.CCv1.2.scaffold159.157	
Blue	01:31,205,460	CB.CBv1.2.scaffold420.23	Pentatricopeptide repeat-containing protein
Blue	01:59,301,385	CC.CCv1.2.scaffold1245.3	
Blue	06:3,821,991	CC.CCv1.2.scaffold523.17	
Blue	06:182,707,804	CC.CCv1.2.scaffold618.15	
Green	06:3,787,311	CA06g01790	
Red	05:608,776	CA05g00490	
*a**	09:1,606,098	CC.CCv1.2.scaffold364.12	
*a**	12:20,136,447	CC.CCv1.2.scaffold505.52	
Blue	12:38,149,033	CC.CCv1.2.scaffold201.26	
Blue	03:101,265,929	CC.CCv1.2.scaffold455.16	
Green	10:124,163,228	CC.CCv1.2.scaffold312.1	
Green	12:65,072,604	CC.CCv1.2.scaffold70.27	
Luminosity	06:3,822,587	CC.CCv1.2.scaffold523.17	
Red	07:5,904,456	CC.CCv1.2.scaffold1259.16	
Hue	06:182,707,823	CC.CCv1.2.scaffold618.15	
Green	06:94,571,346	CC.CCv1.2.scaffold207.3	
Hue	08:13,483,522	CC.CCv1.2.scaffold607.21	
*L**	11:24,777,623	CC.CCv1.2.scaffold375.27	

Table 1 (continued)

Color trait	SNP position	Gene ID	Annotation
Chroma	12:38,149,046	CC.CCv1.2.scaffold201.26	
*L**	02:41,407,729	CC.CCv1.2.scaffold122.13	
Red	11:24,776,180	CC.CCv1.2.scaffold375.27	
Luminosity	02:41,407,693	CC.CCv1.2.scaffold122.13	
*a**	04:16,430,916	CC.CCv1.2.scaffold265.36	Probable LRR receptor-like serine/threonine-protein kinase
Luminosity	05:3,142,405	CC.CCv1.2.scaffold221.3	
Luminosity	04:7,751,133	CC.CCv1.2.scaffold609.15	
Red	04:15,817,504	CB.CBv1.2.scaffold369.41	
Red	04:21,150,437	CC.CCv1.2.scaffold302.13	
Red	12:212,529,651	CC.CCv1.2.scaffold980.7	
Blue	05:3,141,726	CC.CCv1.2.scaffold221.3	
Chroma	04:10,832,611	CB.CBv1.2.scaffold2239.2	
*L**	04:128,593	CB.CBv1.2.scaffold1727.1	
Luminosity	10:216,624,158	CC.CCv1.2.scaffold741.3	
*a**	01:2,876,089	CA04g04460	Putative receptor-like protein kinase
*b**	04:13,533,319	CA04g04460	
*a**	06:12,621,185	CA06g03600	Serine/threonine-protein kinase CDL1
Chroma	11:234,774,398	CB.CBv1.2.scaffold1257.2	
*L**	02:41,407,729	CC.CCv1.2.scaffold122.12	Splicing factor U2af large subunit B isoform X4
Luminosity	02:41,407,693	CC.CCv1.2.scaffold122.12	

cytochrome P450, ABC transporters, and NAC domain-containing protein functions were associated with all *Capsicum* species in all color traits analyzed. Moreover, two genes encoding a flavin-containing monooxygenase were found in the *b** and chroma GWAS for *C. chinense*, while genes with callose synthase 5 protein function were associated with *C. frutescens* in the blue and chroma traits. Genes with calpain-type cysteine protease *DEK1* and receptor-like protein kinase functions were associated with *C. baccatum* and *C. annuum* in the *a**, *b**, and luminosity color traits. Nevertheless, *Capsicum* species exhibit comparable genetic elements contributing to their color-related phenotypic traits.

KEGG analysis using associated genes across the species

The GWAS conducted for color traits in *Capsicum* species revealed several key insights after the associated genes were subjected to pathway analysis. Notably, this analysis identified a total of 27 significantly enriched pathways, as indicated by a *p*-value of ≤ 0.05, and detailed in Fig. 6a of the study. Among these, the main KEGG pathways that stood out were 'metabolic pathways' and 'biosynthesis of secondary metabolites,' both of which contained the highest number of associated genes. Furthermore, additional pathways such as 'plant-pathogen interaction,' 'phenylpropanoid biosynthesis,' 'ABC transporters,' 'mRNA surveillance pathway,' and 'autophagy-other' were also found to be significantly enriched. These findings underscore the complex genetic basis of fruit coloration in *Capsicum* species, highlighting the involvement of a diverse array of biological pathways.

Gene expression of associated genes as revealed by RNA-seq

Furthermore, we analyzed the expression profiles of genes associated with fruit color in *Capsicum chinense*, using their FPKM values from RNA-seq data, as shown in Table S10. An expression heat map visualized the specific gene expression patterns of 25 associated genes in three *C. chinense* accessions — Naga Morich, PI 224448, and PI 257129 — which are classified based on their fruit colors: red, yellow, and brown, respectively. Among these genes, eight encode a pentatricopeptide repeat-containing protein, showing varied expressions across the cultivars. This suggests a pivotal role of PPR genes in determining the fruit color of *C. chinense*. Specifically, four PPR genes were significantly upregulated in Naga Morich (red fruit), while the other four were downregulated. Additionally, a methyltransferase and an LRR

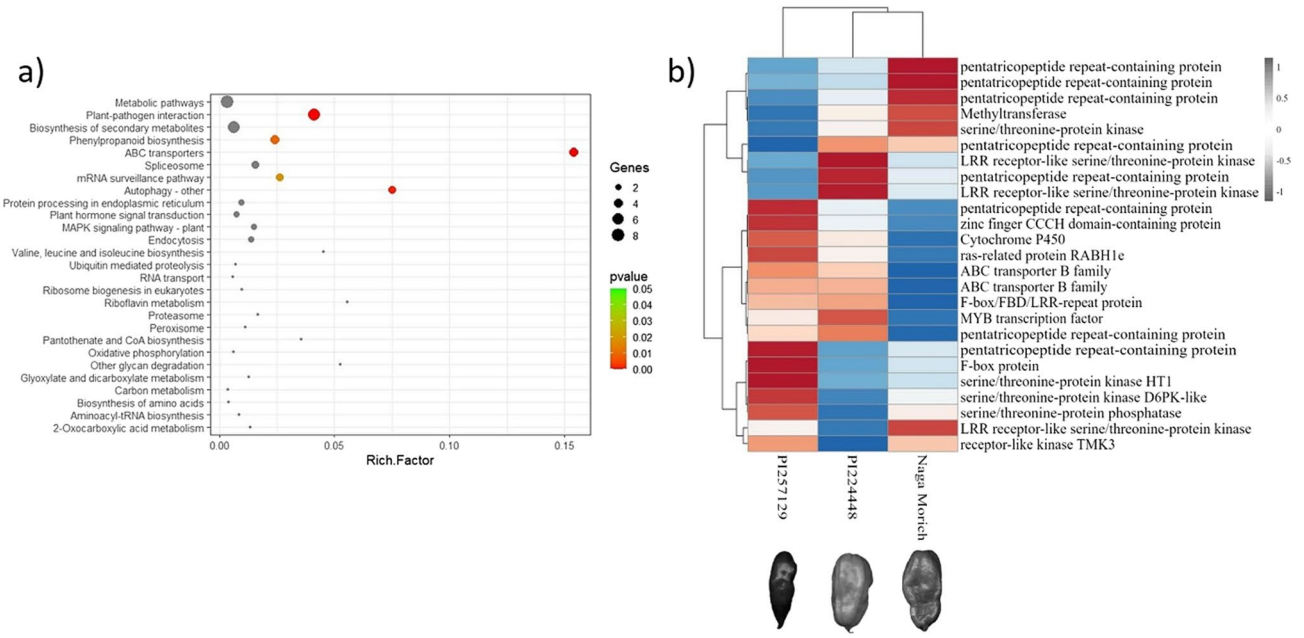


Fig. 6 Pathway enrichment analysis of associated genes for color traits. **a** KEGG pathway enrichment analysis groups genes into various pathways. Level of enrichment can be noted from the scale **b** log2 transformed FPKM of associated genes from the RNAseq from three diverse *C. chinense* accessions (Naga morich, PI 224448, and PI 257129). Level of expressions can be noted on the scale (red, upregulated, and blue, downregulated) (color figure online)

receptor-like serine/threonine-protein kinase exhibited high upregulation in the Naga Morich cultivar compared to those in yellow and brown pepper fruits, as depicted in Fig. 6b. In contrast, a cytochrome P450 and a MYB transcription factor showed high upregulation in PI 257129 (brown fruit) and PI 224448 (yellow fruit), respectively. Furthermore, ABC transporters were upregulated in the brown/yellow fruit accessions, while exhibiting lower expression in the red fruit. Lastly, F-box proteins and serine/threonine-protein kinase/phosphatase were notably upregulated in the brown pepper accession compared to the red and yellow fruits.

KASP assay genotyping for marker validation

One of the notable similarities across the four *Capsicum* species is the recurrence of specific gene types, such as pentatricopeptide repeat-containing proteins, cytochrome P450, and MYB transcription factor. These genes are often associated with significant SNPs for the various traits described earlier. For each *Capsicum* species, we present the allelic effects of the major candidate genes as boxplots in Fig. 7. Selected genes encoding a PPR-containing protein: CB.CBv1.2.scaffold420.23 (01:31,205,460), CC.CCv1.2.scaffold445.33 (06:250,598,054), CC.CCv1.2.scaffold375.27 (11:24,776,180), and CA06g01790 (06:3,787,311); Cytochrome P450: CB.CBv1.2.scaffold473.5 (08:45,351,919), CC.CCv1.2.scaffold289.61 (07:217,590,754), CC.CCv1.2.scaffold58.22 (12:155,572,462), CA12g02930 (12:6,341,820); MYB transcription factor CB.CBv1.2.scaffold484.48 (06:243,279,454), CC.CCv1.2.scaffold82.12 (03:84,522,766), CC.CCv1.2.scaffold73.19 (03:122,791,855), CA03g33530 (03:251,878,883). Our observations indicated that in *C. baccatum,* for the 01:31,205,460 SNP, pepper accessions carrying the major allele AA tended to exhibit a yellow color, while those with the minor allele GG displayed a red hue. Similar color variations, ranging from orange to red, were observed in ripe fruits of the other species as well. Additionally, we developed KASP assays for SNP pairs from the GWAS, specifically targeting two allele-specific primers and a common reverse primer for selected markers 01:31,205,460 and 08:45,351,919, located on PPR-containing protein and Cytochrome P450, respectively (Table S11). These markers effectively resolved three distinct groups: homozygous for allele 1, heterozygous, and homozygous for allele 2. Notably, there was a consistent correlation between the marker genotype and the phenotype with color traits, as illustrated in Fig. 8. These results suggest that markers 01:31,205,460 and 08:45,351,919 could be valuable in future breeding programs for developing high-quality colored pepper varieties. However, further validation of these markers is recommended in all *Capsicum* species.

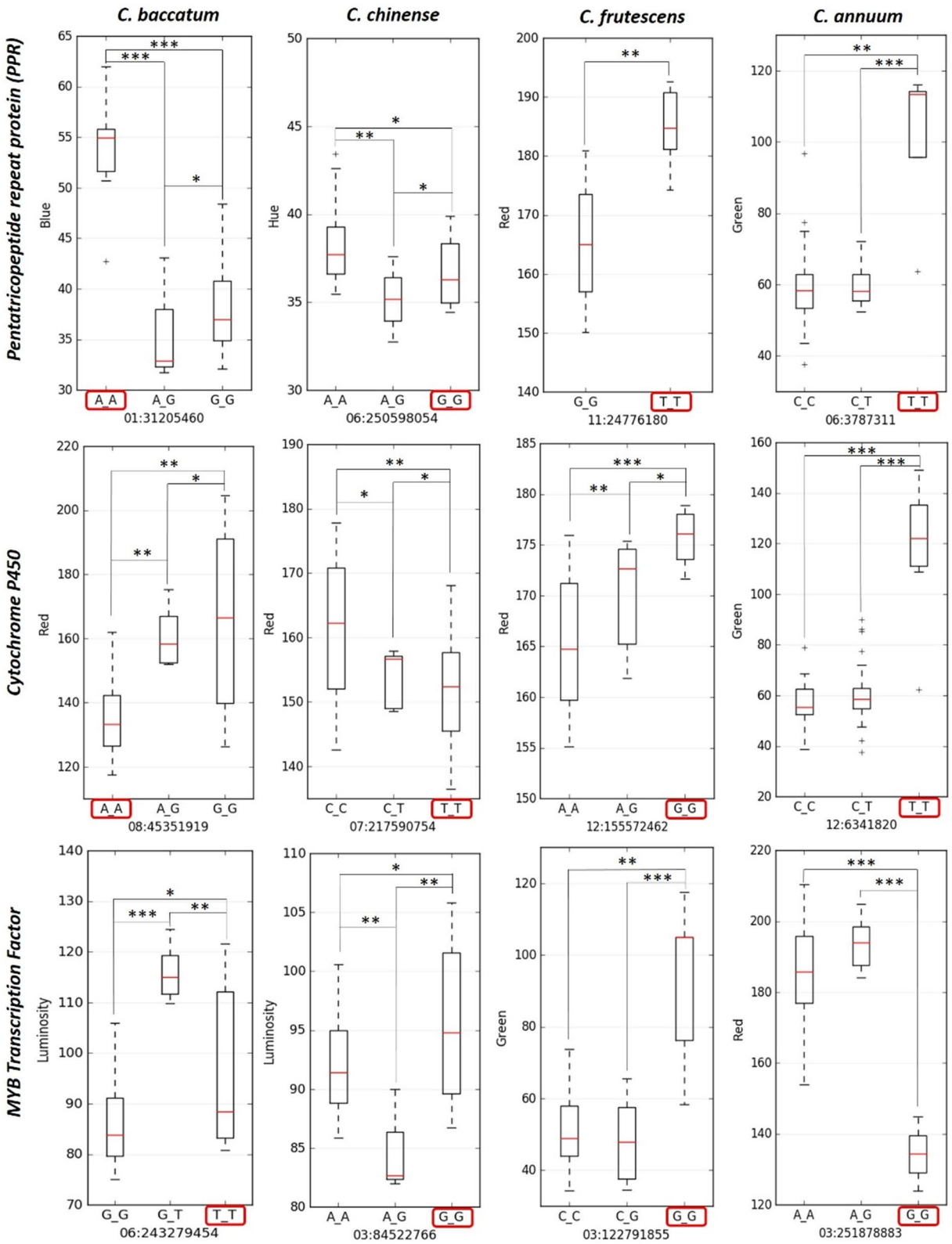


Fig. 7 Allelic effects of significantly associated SNP loci for Pentatricopeptide repeat-containing protein, Cytochrome P450, and MYB transcription factor among *Capsicum* species analyzed. Effects can be noted on the *Y*-axis. Red rectangles represent the minor allele for each SNP. Statistical significance between alleles is indicated as follows: *, $p \leq 0.05$, **, $p \leq 0.01$, ***, $p \leq 0.001$ level (color figure online)

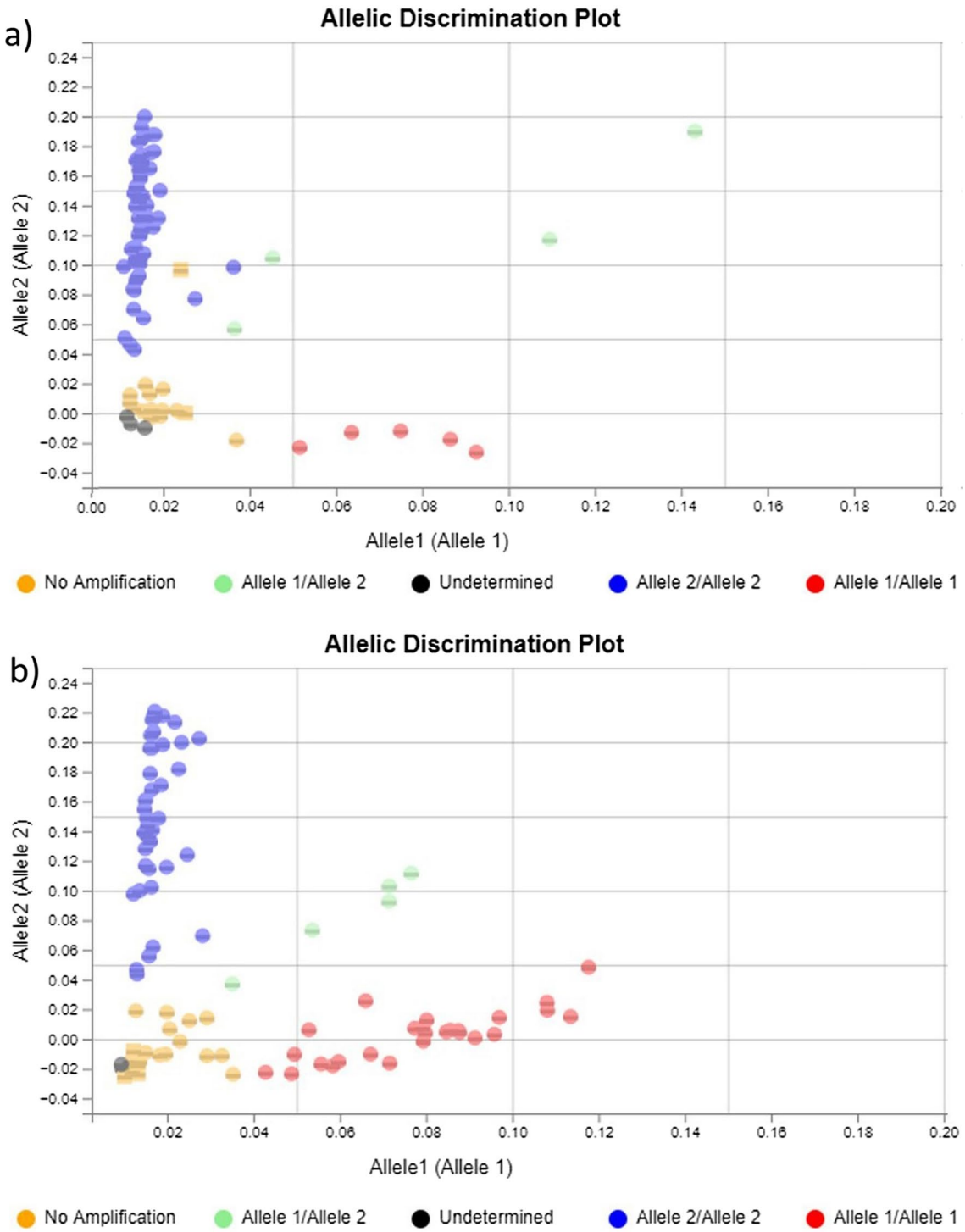


Fig. 8 KASP assays for **a** Chr01:31,205,460 (pentatricopeptide repeat-containing protein) and **b** 08:45,351,919 (cytochrome P450) in the *C. baccatum* collection. Blue and red circles indicate individuals homozygous for the FAM-labeled allele and HEX-labeled allele, respectively. Green circles correspond to heterozygous individuals, and black dots represent the non-template control (NTC) (color figure online)

Discussion

Fruit color is a critical genetic characteristic with significant nutritional and economic implications (Kayesh et al. 2013). It also plays a role in attracting birds, which assists in seed dispersal (Borges 2015). The diversity in the vibrant colors that fruits exhibit upon ripening, both externally and internally, is attributed to natural genetic variation (Han et al. 2018). This study meticulously analyzes color components in 665 accessions from four different *Capsicum* species, identifying a spectrum of genetic elements linked to these variations. The genes identified, such as Cytochrome P450 enzymes, MYB family proteins, Pentatricopeptide repeat-containing proteins (PPRs), ABC transporters, serine/threonine-protein kinases, F-box/kelch-repeat proteins, and LRR receptor-like serine/threonine-protein kinases, appear consistently across the examined species.

These genes are also pivotal in determining the coloration of other agricultural staples such as apples, grapes, and rice (Chen et al. 2021; He et al. 2010; Mbanjo et al. 2020). Additionally, numerous studies have implicated these genes in various molecular mechanisms, particularly those involved in the synthesis, modification, and accumulation of secondary metabolites like flavonoids and carotenoids, highlighting their importance in plant pigmentation (Gonzali and Perata 2021; Kapoor et al. 2022a; Wang et al. 2023). KEGG enrichment analysis of the associated genes revealed several metabolic pathways linked to phenylpropanoid biosynthesis, secondary metabolites, ABC transporters, and plant-pathogen interaction. Notably, there was an observed increase in the activity of phenylalanine ammonia-lyase (*PAL*), a crucial enzyme in the phenylpropanoid pathway (Kim and Hwang 2014). This increased PAL activity is strongly correlated with changes in fruit color and the accumulation of anthocyanin and secondary metabolites (Liu et al. 2023). These findings highlight the significant role of these metabolic pathways and specific enzymes in the biochemical processes affecting fruit pigmentation.

Our GWAS associated seven isoforms of Cytochrome P450 (*CYP450*) enzymes as pivotal in the fruit color regulation of various *Capsicum* species. These enzymes are integral in altering carotenoid molecules, affecting their stability, solubility, and, ultimately, the color of the fruit (Xu et al. 2015). It is established that Cytochrome P450 enzymes also modify flavonoid molecules, influencing the properties of anthocyanins, including their stability and solubility (Tanaka and Brugliera 2013). Beyond their involvement in carotenoids, these enzymes are recognized for their role in the biosynthesis of flavonoids and apocarotenoids (Ahmad et al. 2019; Moreno et al. 2021), as well as in regulating α-carotene levels in carrots (Coe et al. 2021). They are also influential in determining flower colors in roses and carnations (Tanaka and Brugliera 2013). Further, recent transcriptomic analyses have uncovered the role of *CYP450s* in carotenoid metabolism affecting fruit color in melon (Diao et al. 2023). In peppers, fruit color is shaped by the differential accumulation of chlorophyll, carotenoids, anthocyanins, and capsanthin (Wang et al. 2023). CYP450 enzymes work alongside β-carotene hydroxylase to convert β-carotene and α-carotene into zeaxanthin and lutein, leading to the yellow to reddish hues in pepper fruits (Sathasivam and Ki 2018). It is also reported that *CYP89A9*, a variant similar to one we identified in our GWAS (*CYP89A9*like-protein), is active in chlorophyll catabolism during leaf senescence in Arabidopsis (Christ et al. 2013). Interestingly, our KASP assay revealed homozygous allele variations within the *C. baccatum* population for a gene (*CYP72A219*-like protein, CB.CBv1.2.scaffold473.5), suggesting the regulatory role of Cytochrome P450 enzymes in pepper fruit coloration and potential as targets for manipulating carotenoid content in *Capsicum* spp. and beyond.

Furthermore, MYB transcription factors emerged as significant players in regulating fruit color in this GWAS. Their presence in both the red and luminosity color dimensions suggests a regulatory role in controlling critical enzymes like phytoene synthase (*PSY)* and phytoene desaturase (*PDS*), both of which are instrumental in carotenoid synthesis (Borevitz et al. 2000; Shumskaya and Wurtzel 2013; Zhou et al. 2022). They are part of a well-conserved regulatory mechanism involving two specific groups of transcription factors—R2R3 MYB and basic helix-loop-helix (bHLH)—known to control anthocyanin biosynthesis in a variety of species (Borevitz et al. 2000; Kui et al. 2010). Several studies have illuminated the role of MYB transcription factors in regulating anthocyanin production in apples, thereby regulating skin and fruit coloration in apple fruits (An et al. 2018; Chagne et al. 2013; Takos et al. 2006). In a recent study on the Japanese plum (*Prunus salicina*), *PsMYB10* was found to be involved in the regulation of fruit skin color (Fiol et al. 2021). In strawberries, the R2R3 MYB transcription factor *FaMYB1* was found to repress the transcription of anthocyanin-related genes, thereby influencing fruit color (Aharoni et al. 2001; Jin and Martin 1999). Additionally, in tomatoes, a set of highly similar R2R3-MYB transcription factors, including *SlAN2, SlAN2-like, SlANT1*, and *SlANT1-like*, were found to increase anthocyanin content significantly (Kiferle et al. 2015; Meng et al. 2015; Schreiber et al. 2012). In chili pepper, the MYB R2R3 transcription factor reportedly regulated anthocyanin biosynthesis (Borovsky et al. 2004; Stommel et al. 2009). The MYB R2R3 physically interacts with WD40 and bHLH proteins to form a transcriptional activation complex to modulate the expression of late biosynthetic genes, including anthocyanin rhamnosyltransferase therefore controlling anthocyanin biosynthesis and accumulation in pepper (Docimo et al. 2016; Stommel

and Dumm 2015). Similarly, a recent GWAS study comprising genetically diverse *Capsicum* accessions revealed associations of MYB transcription factor encoding genes with the ripe fruit color (McLeod et al. 2023). The diverse roles of MYB transcription factors in modulating the expression of early biosynthetic genes encoding key enzymes in the anthocyanin biosynthesis pathway such as l-phenylalanine ammonia-lyase (*PAL*), cinnamate 4-hydroxylase (*C4H*), and 4-coumarate: CoA ligase (*4CL*) and their transporters make them critical components in the broader landscape of fruit coloration and phytonutrient content in pepper fruits (Venkatesh et al. 2023).

Pentatricopeptide repeat-containing proteins (PPRs) appear in all four GWAS studies in the current research, suggesting a potentially universal role in color-related traits. PPR genes influence organellar mRNA transcripts, regulating the size of chloroplasts and the structure of thylakoid membranes (Wang et al. 2021b; Williams and Barkan 2003). A genomewide study in watermelon identified the causative effects of PPR proteins on flesh color in watermelon (Subburaj et al. 2020). In this study, SNPs present in four PPR genes significantly correlated with the variation in flesh color. In tomato, the *cnr* (color-less nonripening) mutant exhibiting colorless pericarp in the mature fruits displayed reduced expression of a PPR-encoding gene (Eriksson et al. 2004). A QTL analysis in melon (*Cucumis melo*) identified a PPR gene (*CmPPR1*) as the candidate gene for the flesh phenotype, and this locus represents one of the two significant loci that control fruit flesh color in melon (Galpaz et al. 2018). A recent genetic mapping study showed that the C_2 locus controlling yellow flesh color in watermelon is cosegregated with the marker derived from a PPR gene, indicating a vital role of PPR genes in regulating fruit color in diverse species (Park et al. 2023). In the current study, we have identified 12 PRR gene isoforms associated with the color spaces in the study. Additionally, we carried out validation of a specific PPR protein, CB.CBv1.2.scaffold420.23, utilizing KASP markers. This validation enabled us to categorize distinct groups based on allele variations and fruit color. Although there is no clear relationship established between PPR proteins and color rate-limiting enzymes such as phytoene synthase (*PSY1*) and capsanthin-capsorubin synthase (*CCS*) located in plastids (Jang et al. 2020), PPR proteins have been implicated in plastid-to-nucleus retrograde signaling. This signaling is known to affect the expression of genes targeted to plastids, thereby regulating color variation in a quantitative manner (Galpaz et al. 2018; Wang et al. 2021a). Moreover, PPRs are also involved in RNA modifications, including RNA editing, RNA stabilization, and splicing (Qin et al. 2021). Altogether, our findings and prior studies suggest a crucial role of PPR-encoding genes in regulating fruit color in pepper and other vegetable and fruit species.

Among the other GWAS hits, we noted several miscellaneous genes, including ABC transporters, F-Box proteins, and Leucine-rich repeat receptor-like protein kinases (LRR-RLKs). A recent study in apples implicated an ABC transporter, *MdABC17*, in regulating anthocyanin accumulation during fruit ripening. Up-regulation of *MdABC17* increased anthocyanin accumulation in fruits (Xiang et al. 2024). This recent discovery supports our finding that ABC transporters may be involved in transporting flavonoids and carotenoids that determine fruit color in peppers. F-Box proteins regulate phosphorylation-mediated ubiquitination and contain a 50-amino-acid motif facilitating protein–protein interactions. They are implicated in stabilizing key enzymes in the carotenoid pathway (Craig and Tyers 1999; Feder et al. 2015; Schulman et al. 2000). Mutation in an F-Box Domain-Containing Protein is shown to be responsible for the brown hull phenotype in rice by altering the accumulation of flavonoids and anthocyanins (Xu et al. 2016). An F-Box/Kelch-Repeat-containing protein, *SlFKF1* in tomato, was reported to be implicated in regulating lycopene content in tomato (Shibuya et al. 2021).

Additionally, a recent GWAS study using a genetically diverse core collection of *Capsicum* accessions identified an association of an F-Box/Kelch-Repeat protein-encoding gene with the ripe fruit color (McLeod et al. 2023). These findings indicate that F-Box proteins may regulate fruit color in chili peppers. LRR-RLKs are primarily involved in mediating responses to abiotic and biotic stresses, but a study in apricots implicated LRR-RLKs in regulating fruit color (Ferik et al. 2022; Soltabayeva et al. 2022). Our analysis also identified several LRR-RLKs associated with different color attributes in chili peppers. It suggests their well-established role in mediating responses to various stresses may also regulate fruit color. Another notable candidate is the Agamous-like MADS-box protein AGL61, previously characterized in the context of fruit ripening and floral colors (Choudhury et al. 2012). Pectin acetylesterase 8-like (PAE) is another candidate identified in the current GWAS studies; it cleaves the acetyl ester bond from pectin, an essential structural polysaccharide in the primary cell wall (Philippe et al. 2017). PAE is negatively regulated in the shelf life of apples and was also identified as a candidate gene in another study on apple fruit color (Wu et al. 2021). Moreover, we also identified NAC TFs in the GWAS for all *Capsicum* species. NAC TFs have been found to directly target different genes in various pathways, such as chlorophyll degradation carotenoid and anthocyanin accumulation, and regulate color transformation during fruit ripening (Liu et al. 2022). The findings of our study reveal diverse genes and their isoforms associated with various color spaces, suggesting wide-ranging convergence among *Capsicum* species, deepening our understanding of the evolution and diversification of the *Capsicum* genus (Kim et al. 2017; Qin et al. 2014).

Conclusions

This comprehensive study reported the genetic intricacies of color variation within four cultivated species of the *Capsicum* genus. Causal SNPs identified through a precise and systematic approach, shedding light on influential genes such as Cytochrome P450, MYB-related genes, PPR genes, and members of the ABC transporter family. These discoveries unravel the complex molecular tapestry that orchestrates the rich spectrum of hues displayed in Capsicum fruits, ranging from the deepest purples to the brightest reds. The KASP markers characterized in this study can pave the way for marker-assisted selection (MAS) techniques in pepper cultivation. The gene families pinpointed here play significant roles in determining the colors of crops, including apples, grapes, rice, and maize. Furthermore, this study provides a genetic repository that can be tapped into to enhance the aesthetic, nutritional, and commercial value of pepper varieties and other agriculturally important crops, highlighting the potential to influence global food systems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-024-04635-8.

Acknowledgements The authors thank Venkata Abburi and Alicia Talavera-Caro for field studies, technical assistance, and help with the analysis and compiling.

Author contribution statement Conceptualization, P.N., U.K.R., and N.B.; Data curation, A.B. and P. N.; Formal analysis, P.N., and A.B.; Funding acquisition, P.N., D.A., and U.K.R.; Investigation, A.B., B.D. and P.N.; Methodology, A.B., P.NA., K.T., C.L.-O., L.I.-M.; Project administration, P.N., and U.K.R.; Software, U.K.R, A.B., P.NA., S.S.K., and D.A.; Supervision, P.N., U.K.R. and N.B.; Validation, A.B., C.L.-O., L.I.-M.; Visualization, P.N., U.K.R.; Writing—original draft, A.B., C.L.-O., P.N., U.K.R., N.B., M.C., D.A., V.B.; Writing—review and editing, P.N., U.K.R., N.B., M.C., D.A., V.B.; All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by the National Institute of Food and Agriculture USDA-NIFA (grant no. 2019–38821-29064, 2023–38821-39807, and wvax-EA-Padma-2024) and the National Science Foundation: NSF Award Number 2318707.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC (2009) Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. Am J Bot 96:1190–1202

Aharoni A, De Vos CHR, Wein M, Sun ZK, Greco R, Kroon A, Mol JNM, O'Connell AP (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J 28:319–332

Ahmad N, Liu JY, Tian X, Noman M, Jameel A, Yao N, Dong YY, Wang N, Li XW, Wang FW, Liu XM, Li HY (2019) Overexpression of a novel cytochrome P450 promotes flavonoid biosynthesis and osmotic stress tolerance in transgenic arabidopsis. Genes. https://doi.org/10.3390/genes10100756

An JP, Wang XF, Li YY, Song LQ, Zhao LL, You CX, Hao YJ (2018) EIN3-LIKE1, MYB1, and ETHYLENE RESPONSE FACTOR3 act in a regulatory loop that synergistically modulates ethylene biosynthesis and anthocyanin accumulation. Plant Physiol 178:808–823

Azlan A, Sultana S, Huei CS, Razman MR (2022) Antioxidant, anti-obesity, nutritional and other beneficial effects of different chili pepper: a review. Molecules. https://doi.org/10.3390/molecules27030898

Ballester AR, Tikunov Y, Molthoff J, Grandillo S, Viquez-Zamora M, de Vos R, de Maagd RA, van Heusden S, Bovy AG (2016) Identification of loci affecting accumulation of secondary metabolites in tomato fruit of a *Solanum lycopersicum* × Solanum chmielewskii introgression line population. Front Plant Sci. https://doi.org/10.3389/fpls.2016.01428

Borevitz JO, Xia YJ, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383–2393

Borges RM (2015) Fruit and seed volatiles: multiple stage settings, actors and props in an evolutionary play. J Indian Inst Sci 95:93–104

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

Borovsky Y, Paran I (2008) Chlorophyll breakdown during pepper fruit ripening in the chlorophyll retainer mutation is impaired at the homolog of the senescence-inducible stay-green gene. Theor Appl Genet 117:235–240

Carrizo García C, Barfuss MH, Sehr EM, Barboza GE, Samuel R, Moscone EA, Ehrendorfer F (2016) Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, *Solanaceae*). Ann Bot 118:35–51

Chagne D, Kui LW, Espley RV, Volz RK, How NM, Rouse S, Brendolise C, Carlisle CM, Kumar S, De Silva N, Micheletti D, McGhie T, Crowhurst RN, Storey RD, Velasco R, Hellens RP, Gardiner SE, Allan AC (2013) An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiol 161:225–239

Chen H, Ji H, Zhu S, Zhu K, Ye J, Deng X (2023) Carotenoid and transcriptome profiles of a novel citrus cultivar 'Jinlegan' reveal mechanisms of yellowish fruit formation. Hortic Adv 1:5

Chen ZJ, Yu L, Liu WJ, Zhang J, Wang N, Chen XS (2021) Research progress of fruit color development in apple (*Malus domestica* Borkh.). Plant Physiol Biochem 162:267–279

Choudhury SR, Roy S, Nag A, Singh SK, Sengupta DN (2012) Characterization of an AGAMOUS-like MADS box protein, a probable constituent of flowering and fruit ripening regulatory system in banana. PLoS ONE. https://doi.org/10.1371/journal.pone.0044361

Christ B, Sussenbacher I, Moser S, Bichsel N, Egert A, Muller T, Krautler B, Hortensteiner S (2013) Cytochrome P450 CYP89A9 is involved in the formation of major chlorophyll catabolites during leaf senescence in arabidopsis. Plant Cell 25:1868–1880

Cingolani P (2022) Variant annotation and functional prediction: SnpEff. In: Ng C, Piscuoglio S (eds) Variant calling: methods and protocols. Springer US, New York, NY, pp 289–314

Coe KM, Ellison S, Senalik D, Dawson J, Simon P (2021) The influence of the Or and carotene hydroxylase genes on carotenoid accumulation in orange carrots *Daucus carota* (L.). Theor Appl Genet 134:3351–3362

Craig KL, Tyers M (1999) The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog Biophys Mol Biol 72:299–328

Darrigues A, Hall J, van der Knaap E, Francis DM, Dujmovic N, Gray S (2008) Tomato analyzer-color test: a new tool for efficient digital phenotyping. J Am Soc Hortic Sci 133:579–586

Deepa N, Kaur C, George B, Singh B, Kapoor HC (2007) Antioxidant constituents in some sweet pepper (*Capsicum annuum* L.) genotypes during maturity. Lwt-Food Sci Technol 40:121–129

Diao QN, Tian SB, Cao YY, Yao DW, Fan HW, Zhang YP (2023) Transcriptome analysis reveals association of carotenoid metabolism pathway with fruit color in melon. Sci Rep. https://doi.org/10.1038/s41598-023-31432-y

Docimo T, Francese G, Ruggiero A, Batelli G, De Palma M, Bassolino L, Toppino L, Rotino GL, Mennella G, Tucci M (2016) Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. Front Plant Sci 6:1233

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:e19379

Eriksson EM, Bovy A, Manning K, Harrison L, Andrews J, De Silva J, Tucker GA, Seymour GB (2004) Effect of the colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. Plant Physiol 136:4184–4197

Feder A, Burger J, Gao S, Lewinsohn E, Katzir N, Schaffer AA, Meir A, Davidovich-Rikanati R, Portnoy V, Gal-On A, Fei ZJ, Kashi Y, Tadmor Y (2015) A kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in muskmelon. Plant Physiol 169:1714–1726

Ferik F, Ates D, Ercisli S, Erdogan A, Orhan E, Tanyolac MB (2022) Genome-wide association links candidate genes to fruit firmness, fruit flesh color, flowering time, and soluble solid content in apricot (*Prunus armeniaca* L.). Mol Biol Rep 49:5283–5291

Fiol A, Garcia-Gomez BE, Jurado-Ruiz F, Alexiou K, Howad W, Aranzana MJ (2021) Characterization of japanese plum (*Prunus salicina*) PsMYB10 alleles reveals structural variation and polymorphisms correlating with fruit skin color. Front Plant Sci. https://doi.org/10.3389/fpls.2021.655267

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V, This P (2009) Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp sativa) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. Genetics 183:1127–1139

Frank CA, Nelson RG, Simonne EH, Behe BK, Simonne AH (2001) Consumer preferences for color, price, and vitamin C content of bell peppers. HortScience 36:795–800

Galpaz N, Gonda I, Shem-Tov D, Barad O, Tzuri G, Lev S, Fei Z, Xu Y, Mao L, Jiao C (2018) Deciphering genetic factors that determine melon fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping. Plant J 94:169–191

Gonzali S, Perata P (2021) Fruit colour and novel mechanisms of genetic regulation of pigment production in tomato fruits. Horticulturae. https://doi.org/10.3390/horticulturae7080259

Han Z, Hu Y, Lv Y, Rose JK, Sun Y, Shen F, Wang Y, Zhang X, Xu X, Wu T (2018) Natural variation underlies differences in ETHYLENE RESPONSE FACTOR17 activity in fruit peel degreening. Plant Physiol 176:2292–2304

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. In: Fleury D, Whitford R (eds) Crop breeding: methods and protocols. Springer New York, pp 75–86

He F, Mu L, Yan GL, Liang NN, Pan QH, Wang J, Reeves MJ, Duan CQ (2010) Biosynthesis of anthocyanins and their regulation in colored grapes. Molecules 15:9057–9091

Howard LR, Talcott ST, Brenes CH, Villalon B (2000) Changes in phytochemical and antioxidant activity of selected pepper cultivars (*Capsicum* species) as influenced by maturity. J Agric Food Chem 48:1713–1720

Hurtado-Hernandez H, Smith PG (1985) Inheritance of mature fruit color in *Capsicum-ANNUUM*-L. J Hered 76:211–213

Jang S-J, Jeong H-B, Jung A, Kang M-Y, Kim S, Ha S-H, Kwon J-K, Kang B-C (2020) Phytoene synthase 2 can compensate for the absence of PSY1 in the control of color in *Capsicum* fruit. J Exp Bot 71:3417–3427

Jin HL, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol 41:577–585

Kapoor L, Simkin AJ, Doss CGP, Siva R (2022a) Fruit ripening: dynamics and integrated analysis of carotenoids and anthocyanins. Bmc Plant Biol. https://doi.org/10.1186/s12870-021-03411-w

Kapoor L, Simkin AJ, Doss CGP, Siva R (2022b) Fruit ripening: dynamics and integrated analysis of carotenoids and anthocyanins. BMC Plant Biol 22:1–22

Kayesh E, Shangguan L, Korir NK, Sun X, Bilkish N, Zhang Y, Han J, Song C, Cheng Z-M, Fang J (2013) Fruit skin color and the role of anthocyanin. Acta Physiol Plant 35:2879–2890

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. https://doi.org/10.1371/journal.pone.0136365

Kim DS, Hwang BK (2014) An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens. J Exp Bot 65:2295–2306

Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, Kim SB, Han K, Lee J, Park M, Lee HA, Lee HY, Lee Y, Oh S, Lee JH, Choi E, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee SC, Kwon JK, Koo N, Hong Y, Kim RW, Kang WH, Huh JH, Kang BC, Yang TJ, Lee YH, Bennetzen JL, Choi D (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. https://doi.org/10.1186/s13059-017-1341-9

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JKC, Sorensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270

Kraft KH, Brown CH, Nabhan GP, Luedeling E, Luna Ruiz JdJ, Coppens d'Eeckenbrugge G, Hijmans RJ, Gepts P (2014) Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. Proc Natl Acad Sci 111:6165–6170

Kui LW, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. Bmc Plant Biol. https://doi.org/10.1186/1471-2229-10-50

Liu G-S, Li H-L, Grierson D, Fu D-Q (2022) NAC transcription factor family regulation of fruit ripening and quality: a review. Cells 11:525

Liu S, Zhang Q, Kollie L, Dong J, Liang Z (2023) Molecular networks of secondary metabolism accumulation in plants: current understanding and future challenges. Ind Crops Prod 201:116901

Mbanjo EGN, Kretzschmar T, Jones H, Ereful N, Blanchard C, Boyd LA, Sreenivasulu N (2020) The genetic basis and nutritional benefits of pigmented rice grain. Front Genet. https://doi.org/10.3389/fgene.2020.00229

McLeod L, Barchi L, Tumino G, Tripodi P, Salinier J, Gros C, Boyaci HF, Ozalp R, Borovsky Y, Schafleitner R, Barchenger D, Finkers R, Brouwer M, Stein N, Rabanus-Wallace MT, Giuliano G, Voorrips R, Paran I, Lefebvre V (2023) Multi-environment association study highlights candidate genes for robust agronomic quantitative trait loci in a novel worldwide *Capsicum* core collection. Plant J. https://doi.org/10.1111/tpj.16425

Mendes ND, Goncalves E (2020) The role of bioactive components found in peppers. Trends Food Sci Technol 99:229–243

Meng X, Yang DY, Li XD, Zhao SY, Sui N, Meng QW (2015) Physiological changes in fruit ripening caused by overexpression of tomato SlAN2, an R2R3-MYB factor. Plant Physiol Biochem 89:24–30

Moreno JC, Mi JN, Alagoz Y, Al-Babili S (2021) Plant apocarotenoids: from retrograde signaling to interspecific communication. Plant J 105:351–375

Munshell AH (1912) A pigment color system and notation. Am J Psychol 23:236–244

Nabi BG, Mukhtar K, Ahmed W, Manzoor MF, Ranjha MMAN, Kieliszek M, Bhat ZF, Aadil RM (2023) Natural pigments: anthocyanins, carotenoids, chlorophylls, and betalains as food colorants in food products. Food Biosci 52:102403

Nankar AN, Tringovska I, Grozeva S, Ganeva D, Kostova D (2020a) Tomato phenotypic diversity determined by combined approaches of conventional and high-throughput tomato analyzer phenotyping. Plants-Basel. https://doi.org/10.3390/plants9020197

Nankar AN, Tringovska I, Grozeva S, Todorova V, Kostova D (2020b) Application of high-throughput phenotyping tool Tomato Analyzer to characterize Balkan Capsicum fruit diversity. Sci Hortic. https://doi.org/10.1016/j.scienta.2019.108862

Natarajan P, Akinmoju TA, Nimmakayala P, Lopez-Ortiz C, Garcia-Lozano M, Thompson BJ, Stommel J, Reddy UK (2020) Integrated metabolomic and transcriptomic analysis to characterize cutin biosynthesis between low-and high-cutin genotypes of *Capsicum* chinense Jacq. Int J Mol Sci 21:1397

Nimmakayala P, Lopez-Ortiz C, Shahi B, Abburi VL, Natarajan P, Kshetry AO, Shinde S, Davenport B, Stommel J, Reddy UK (2021) Exploration into natural variation for genes associated with fruit shape and size among *Capsicum* chinense collections. Genomics 113:3002–3014

Park G, Shahwar D, Jang G, Shin J, Kwon G, Kim Y, Hong CO, Jin B, Kim H, Lee O, Park Y (2023) Identification of a novel locus C2 controlling canary yellow flesh color in watermelons. Front Genet. https://doi.org/10.3389/fgene.2023.1256627

Philippe F, Pelloux J, Rayon C (2017) Plant pectin acetylesterase structure and function: new insights from bioinformatic analysis. BMC Genomics. https://doi.org/10.1186/s12864-017-3833-0

Pickersgill B (1971) Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). Evolution 25:683–691

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, Gonzalez-Morales S, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernandez S, Leyva-Gonzalez MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci USA 111:5135–5140

Qin TF, Zhao P, Sun JL, Zhao YP, Zhang YX, Yang QY, Wang WP, Chen ZQ, Mai TF, Zou YY, Liu GX, Hao W (2021) Research progress of PPR proteins in RNA editing, stress response, plant growth and development. Front Genet. https://doi.org/10.3389/fgene.2021.765580

Robarts D, Wolfe A, Jourdan P (2012) Relationships among germplasm accessions of viola and phlox examined by analysis of flower color and morphology using tomato analyzer (TM) image analysis software. HortScience 47:S344–S344

Rodriguez-Uribe L, Guzman I, Rajapakse W, Richins RD, O'Connell MA (2012) Carotenoid accumulation in orange-pigmented *Capsicum annuum* fruit, regulated at multiple levels. J Exp Bot 63:517–526

Sathasivam R, Ki J-S (2018) A review of the biological activities of microalgal carotenoids and their potential use in healthcare and cosmetic industries. Mar Drugs 16:26

Schifferstein HNJ, Wehrle T, Carbon CC (2019) Consumer expectations for vegetables with typical and atypical colors: the case of carrots. Food Qual Prefer 72:98–108

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

Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ERE, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavietich NP (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 408:381–386

Scossa F, Roda F, Tohge T, Georgiev MI, Fernie AR (2019) The hot and the colorful: understanding the metabolism, genetics and evolution of consumer preferred metabolic traits in pepper and related species. Crit Rev Plant Sci 38:339–381

Shibuya T, Nishiyama M, Kato K, Kanayama Y (2021) Characterization of the FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 Homolog SlFKF1 in tomato as a model for plants with fleshy fruit. Int J Mol Sci. https://doi.org/10.3390/ijms22041735

Shumskaya M, Wurtzel ET (2013) The carotenoid biosynthetic pathway: thinking in all dimensions. Plant Sci 208:58–63

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE 7:e40563

Soltabayeva A, Dauletova N, Serik S, Sandybek M, Omondi JO, Kurmanbayeva A, Srivastava S (2022) Receptor-like kinases (LRR-RLKs) in response of plants to biotic and abiotic stresses. Plants-Basel. https://doi.org/10.3390/plants11192660

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

Stommel JR, Lightbourn GJ, Winkel BS, Griesbach RJ (2009) transcription factor families regulate the anthocyanin biosynthetic pathway in *Capsicum annuum*. J Am Soc Hortic Sci 134:244–251

Subburaj S, Tu LH, Lee K, Park GS, Lee H, Chun JP, Lim YP, Park MW, McGregor C, Lee GJ (2020) A genome-wide analysis of the pentatricopeptide repeat (PPR) gene family and ppr-derived markers for flesh color in watermelon (*Citrullus lanatus*). Genes. https://doi.org/10.3390/genes11101125

Takos AM, Jaffe FW, Jacob SR, Bogs J, Robinson SP, Walker AR (2006) Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol 142:1216–1232

Tanaka Y, Brugliera F (2013) Flower colour and cytochromes P450. Philos Trans R Soc B-Biol Sci. https://doi.org/10.1098/rstb.2012.0432

Tong H, Nankar AN, Liu J, Todorova V, Ganeva D, Grozeva S, Tringovska I, Pasev G, Radeva-Ivanova V, Gechev T (2022) Genomic prediction of morphometric and colorimetric traits in solanaceous fruits. Hortic Res. https://doi.org/10.1093/hr/uhac072

Venkatesh J, Lee S-Y, Back S, Kim T-G, Kim GW, Kim J-M, Kwon J-K, Kang B-C (2023) Update on the genetic and molecular regulation of the biosynthetic pathways underlying pepper fruit color and pungency. Curr Plant Biol 35–36:100303

Villa-Rivera MG, Ochoa-Alejo N (2021) Transcriptional regulation of ripening in chili pepper fruits (*Capsicum* spp.). Int J Mol Sci. https://doi.org/10.3390/ijms222212151

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 76:783–793

Walsh BM, Hoot SB (2001) Phylogenetic relationships of *Capsicum* (*Solanaceae*) using DNA sequences from two noncoding regions: the chloroplast atpB-rbcL spacer region and nuclear waxy introns. Int J Plant Sci 162:1409–1418

Wang L, Zhong Y, Liu J, Ma R, Miao Y, Chen W, Zheng J, Pang X, Wan H (2023) Pigment biosynthesis and molecular genetics of fruit color in pepper. Plants (basel). https://doi.org/10.3390/plants12112156

Wang X, An Y, Xu P, Xiao J (2021a) Functioning of PPR proteins in organelle RNA metabolism and chloroplast biogenesis. Front Plant Sci 12:627501

Wang XW, An YQ, Xu P, Xiao JW (2021b) Functioning of PPR proteins in organelle RNA metabolism and chloroplast biogenesis. Front Plant Sci. https://doi.org/10.3389/fpls.2021.627501

Weale ME (2010) Quality control for genome-wide association studies. In: Barnes MR, Breen G (eds) Genetic variation: methods and protocols. Springer, pp 341–372

Williams PM, Barkan A (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. Plant J 36:675–686

Wu B, Shen F, Chen CJ, Liu L, Wang X, Zheng WY, Deng Y, Wang T, Huang ZY, Xiao C, Zhou Q, Wang Y, Wu T, Xu XF, Han ZH, Zhang XZ (2021) Natural variations in a pectin acetylesterase gene, MdPAE10, contribute to prolonged apple fruit shelf life. Plant Genome. https://doi.org/10.1002/tpg2.20084

Xiang Y, Yu Huang X, Zhao Y-W, Wang C-K, Sun Q, Hu D-G (2024) Role of an ATP binding cassette (ABC) transporter MdABCI17 in the anthocyanin accumulation of apple. Sci Hortic 323:112502

Xu J, Wang XY, Guo WZ (2015) The cytochrome P450 superfamily: key players in plant development and defense. J Integr Agric 14:1673–1686

Xu X, Zhang XB, Shi YF, Wang HM, Feng BH, Li XH, Huang QN, Song LX, Guo D, He Y, Wu JL (2016) A point mutation in an F-box domain-containing protein is responsible for brown hull phenotype in rice. Rice Sci 23:1–8

Zhou XS, Rao SM, Wrightstone E, Sun TH, Lui ACW, Welsch R, Li L (2022) Phytoene synthase: the key rate-limiting enzyme of carotenoid biosynthesis in plants. Front Plant Sci. https://doi.org/10.3389/fpls.2022.884720