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.

Genetic Tapestry of Capsicum Fruit Colors: A Comparative Analysis of Four Cultivated Species

Ambika Bhattarai
West Virginia State University
Padma Nimmakayala
West Virginia State University
Brittany Davenport
West Virginia State University
Purushothaman Natarajan
West Virginia State University
Alicia Talavera-Caro
West Virginia State University
Carlos Lopez-Ortiz
West Virginia State University
Lizbeth Ibarra-Muñoz
Universidad Autonoma de Coahuila - Unidad Torreon
Manohar Chakrabarti
The University of Texas Rio Grande Valley
Vagner Benedito
West Virginia University
Nagamani Balagurusamy
Universidad Autonoma de Coahuila - Unidad Torreon
Umesh K Reddy
ureddy@wvstateu.edu

West Virginia State University https://orcid.org/0000-0003-4611-7143

Research Article

Keywords: Capsicum, pepper fruit, color diversity, GWAS

Posted Date: December 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3580934/v1

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

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on May 14th, 2024. See the published version at https://doi.org/10.1007/s00122-024-04635-8.

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 content, 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 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 (Chr01:31205460) and a cytochrome P450 enzyme (Chr08:45351919) 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.

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 available at the Pepper Genome Database (http://peppergenome.snu.ac.kr) using BWA-MEM, as described in (Kim et al. 2017; Kim et al. 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.

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; Nankar et al. 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 Munsell 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 (detailed definitions of fruit color measurements are given in Table S2). 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 Mapping

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 scrutinized 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, available in Appendix, Figure S1. We established a significance threshold based on false discovery rate (FDR) assessments.

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 Single Nucleotide Polymorphisms (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 *C. annuum, C. baccatum, C. frutescens*, and *C. chinense*. 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.

Kompetitive allele-specific PCR (KASP)

Sequences flanking selected Single Nucleotide Polymorphisms (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 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 minutes, followed by ten cycles of denaturation at 94°C for 20 seconds, and annealing/extension from 61°C to 55°C decreasing by 0.6°C per cycle for 60 seconds each, and concluding with 35 cycles at 94°C for 20 seconds and a constant 55°C for 60 seconds for annealing/extension. The genotyping results were analyzed using the StepOne plus auto-caller software from Applied Biosystems.

Results

SNPs development and phenotyping for fruit color

Utilizing the GB-easy Pipeline, we generated a robust set of Single Nucleotide Polymorphisms (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 a staggering 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.

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 nine color spaces—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 S3-S6). 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 spaces within our *Capsicum* accessions, representing all four species under study (Fig. 1 and Figure S1). This diversity is vividly illustrated in box plots presented in Fig. 2. In addition to the box plots, the frequency distribution of these nine color spaces was analyzed and depicted in Figure S2. The data exhibit a 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 scrutinize 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 S3. The analysis revealed intriguing patterns of associations between color spaces, some of which were species-specific. For instance, within the *C. baccatum* accessions, nearly all color spaces displayed positive correlations except for color space a*. This metric showed a pronounced negative correlation with green and hue while maintaining weak associations with other color spaces. In contrast, the *C. chinense* accessions revealed that color space a* had strong negative correlations with green, hue, and L*. Moreover, the color space 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 space showed negligible to weak correlations with the remaining color spaces. In the *C. annuum* collections, the color space a* was negatively correlated with green, L*, blue, and hue. Significantly, a* trait positively correlated with chroma in this species. Furthermore, color space b* in *C. annuum* presented an intriguing divergence, showing little to no correlation with the other color spaces. 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 Structure Analyses

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 Single Nucleotide Polymorphisms (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 Fig. 3, 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. 4). As illustrated in numeric PCA, these results unambiguously demonstrate that the chosen color spaces 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 spaces: Red, Green, Blue, Luminosity, L*, a*, b*, Hue, and Chroma. The resulting SNP-trait associations were elegantly visualized through Manhattan plots, tailored for each color dimension. Our rigorous statistical framework identified many significantly associated loci, characterized by -log10 P-value exceeding 3. 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 whittled the list to 40 pivotal associations for *C. annuum*, 35 for C. *baccatum*, 61 for *C. chinense*, and 42 for *C. frutescens.* Manhattan plots showing the associated SNPs for each color trait and *Capsicum* species are presented in the supplementary file 1. A comprehensive tabulation of the 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 elaborated in Table S7. 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 S8.

Red*

The comprehensive genetic analysis of four *Capsicum* species—*C. baccatum, C. chinense, C. frutescens*, and *C. annuum*—revealed a variety of significant Single Nucleotide Polymorphisms (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 a staggering 52%, with an SNP at position 11:24776180 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 03:251878883, linked to a MYB-related protein, accounted for 25% of the observed variance.

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 228959560 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 scrutiny were identified on chromosomes 1, 6, 11, and 12, accounting for a variance from 8–13% in phenotype. An SNP at the 3787311 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 01:31205460 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 11–17% variance. SNPs on chromosome positions 03:6730447 and 10:28733075 were

significantly associated with F-box/kelch-repeat proteins. In the case of *C. frutescens*, seven pivotal SNPs were located on chromosomes 2, 3, 6, 8, and 12, and they accounted for phenotypic variance from 9–13%. These SNPs were recurrently linked to pentatricopeptide repeat-containing proteins, especially in the blue color space. Finally, for *C. annuum*, three SNPs that passed the FDR correction were located on chromosomes 3, 5, and 12, explaining the phenotypic variance of 10–14%. Specifically, the SNP at position 03:24306417 on chromosome 3 was associated with a protein FAR1-RELATED SEQUENCE 11-like gene and was particularly noteworthy for blue color space variance.

Luminosity*

For *C. baccatum*, four standout SNPs across chromosomes 2, 4, and 6 explained phenotypic variance between 8% and 21% (Fig. 5). Notably, the SNP on chromosome 04:8593005 was linked to a calpain-type cysteine protease, whereas the SNP at 02:148766979 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 03:84522766 was linked with an arginine/serine-rich coiled-coil protein 2 isoform X2 and a MYB-like protein I. Additionally, SNPs on chromosomes 04:7751133, 05:3142405, and 10:216624158 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 02:41407693 and 06:3822587 explaining 11% and 12% phenotypic variance, respectively. Notably, SNPs at chromosome 06:230103252 and 06:230103244 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 01:169031114, 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 12:39093634 was related to a calcium-dependent protein kinase, while another SNP at 12:228886503 was linked to an ABC transporter B family member. An additional SNP at 04:128593 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 01:153341594 and 01:98936335, 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 06:210658508 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 01:252255946 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 03:279762481 was linked to a probable LRR receptor-like serine/threonine-protein kinase, while the one at 07:210731968 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 11–12%. For example, the SNP at position 09:133866441 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 03:135620678 was associated with peroxidase 5-like, explicitly affecting the color space 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 08:144332220 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 a staggering 54%. For example, the SNP at chromosome 05:217136990 was a cofactor for b* variance associated with ras-related and BTB/POZ domain-containing proteins. Another SNP at chromosome 04:224586097, 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 01:53151302, 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 04:13533319 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 04:15303598 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 08:193074112 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 07:211354155, associated with an F-box protein, and 12:224304164, 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 12:234434115 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 space.

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 04:10832611, linked to a probable LRR receptor-like serine/threonine-protein kinase, and 08:45351919, 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 10–21%. Notable SNPs included those at chromosome positions 02:132385192 and 01:53651476, 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 10–62%. For example, an SNP at 03:243301602 was associated with a nuclear transcription factor, while another at 06:19381223 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 12:232533746 and 12:1969418 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. 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 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.

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

Color Trait	SNP position	Gene ID	Annotation
Green	01:42272375	CC.CCv1.2.scaffold549.1	ABC transporter A family
Chroma	12:232533746	CA12g20900	ABC transporter B family
Hue	06:2634	CC.CCv1.2.scaffold1019.1	
L*	12:228886503	CB.CBv1.2.scaffold2474.1	
Blue	02:132022079	CC.CCv1.2.scaffold823.48	
Chroma	08:26089056	CC.CCv1.2.scaffold183.46	
a*	04:20870037	CC.CCv1.2.scaffold302.16	ABC transporter G family
Luminosity	06:230103252	CC.CCv1.2.scaffold31.72	
Red	05:1435954	CC.CCv1.2.scaffold856.8	ABC transporter I family
Blue	12:38149033	CC.CCv1.2.scaffold201.30	callose synthase 5
Chroma	12:38149046	CC.CCv1.2.scaffold201.30	
b*	04:8335730	CB.CBv1.2.scaffold1012.15	calpain-type cysteine protease DEK1
Luminosity	04:8593005	CB.CBv1.2.scaffold1012.15	
Green	12:6341820	CA12g02930	cytochrome P450
Red	12:155572462	CC.CCv1.2.scaffold58.22	
	08:45351919	CB.CBv1.2.scaffold473.5	
Blue	03:6730447	CC.CCv1.2.scaffold647.51	F-box/kelch-repeat protein At3g23880-like
Blue	03:6730447	CC.CCv1.2.scaffold647.52	
Blue	10:28733075	CC.CCv1.2.scaffold97.1	
Chroma	12:1969418	CA12g00600	
Hue	03:17947103	CA03g06280	
L*	10:10391356	CC.CCv1.2.scaffold1234.4	
L*	12:226262158	CC.CCv1.2.scaffold710.69	
Luminosity	12:224920669	CC.CCv1.2.scaffold1389.1	
b*	06:192909983	CC.CCv1.2.scaffold501.1	flavin-containing monooxygenase FMO GS-OX-like 4
Chroma	06:192759203	CC.CCv1.2.scaffold501.5	

Color Trait	SNP position	Gene ID	Annotation
b*	06:205140356	CA06g14960	NAC domain-containing protein 78-like
Chroma	12:33932768	CB.CBv1.2.scaffold2291.2	
Chroma	05:233972444	CC.CCv1.2.scaffold159.157	
Blue	01:31205460	CB.CBv1.2.scaffold420.23	pentatricopeptide repeat-containing protein
Blue	01:59301385	CC.CCv1.2.scaffold1245.3	
Blue	06:3821991	CC.CCv1.2.scaffold523.17	
Blue	06:182707804	CC.CCv1.2.scaffold618.15	
Green	06:3787311	CA06g01790	
Red	05:608776	CA05g00490	
a*	09:1606098	CC.CCv1.2.scaffold364.12	
a*	12:20136447	CC.CCv1.2.scaffold505.52	
Blue	12:38149033	CC.CCv1.2.scaffold201.26	
Blue	03:101265929	CC.CCv1.2.scaffold455.16	
Green	10:124163228	CC.CCv1.2.scaffold312.1	
Green	12:65072604	CC.CCv1.2.scaffold70.27	
Luminosity	06:3822587	CC.CCv1.2.scaffold523.17	
Red	07:5904456	CC.CCv1.2.scaffold1259.16	
Hue	06:182707823	CC.CCv1.2.scaffold618.15	
Green	06:94571346	CC.CCv1.2.scaffold207.3	
Hue	08:13483522	CC.CCv1.2.scaffold607.21	
L*	11:24777623	CC.CCv1.2.scaffold375.27	
Chroma	12:38149046	CC.CCv1.2.scaffold201.26	
L*	02:41407729	CC.CCv1.2.scaffold122.13	
Red	11:24776180	CC.CCv1.2.scaffold375.27	
Luminosity	02:41407693	CC.CCv1.2.scaffold122.13	
a*	04:16430916	CC.CCv1.2.scaffold265.36	probable LRR receptor-like serine/threonine-protein kinase
Luminosity	05:3142405	CC.CCv1.2.scaffold221.3	
Luminosity	04:7751133	CC.CCv1.2.scaffold609.15	

Color Trait	SNP position	Gene ID	Annotation
Red	04:15817504	CB.CBv1.2.scaffold369.41	
Red	04:21150437	CC.CCv1.2.scaffold302.13	
Red	12:212529651	CC.CCv1.2.scaffold980.7	
Blue	05:3141726	CC.CCv1.2.scaffold221.3	
Chroma	04:10832611	CB.CBv1.2.scaffold2239.2	
L*	04:128593	CB.CBv1.2.scaffold1727.1	
Luminosity	10:216624158	CC.CCv1.2.scaffold741.3	
a*	01:2876089	CA04g04460	putative receptor-like protein kinase
b*	04:13533319	CA04g04460	
a*	06:12621185	CA06g03600	serine/threonine-protein kinase CDL1
Chroma	11:234774398	CB.CBv1.2.scaffold1257.2	
L*	02:41407729	CC.CCv1.2.scaffold122.12	splicing factor U2af large subunit B isoform X4
Luminosity	02:41407693	CC.CCv1.2.scaffold122.12	

KASP assay genotyping for marker validation

One of the striking similarities across the four *Capsicum* species is the recurrence of certain types of genes, such as pentatricopeptide repeat-containing proteins and Cytochrome P450, which are often associated with significant SNPs for various traits described above. We selected a gene for each *Capsicum* species, CB.CBv1.2.scaffold420.23 (01:31205460), CC.CCv1.2.scaffold1081.1 (07:134725761), CA05g00490 (05:608776) encoding a PPR-containing protein to determine their allelic effects are presented as boxplots in Fig. 6. Our observations indicated that in *C. baccatum*, pepper accessions carrying the major allele AA exhibited a yellow color, whereas mature fruits with the minor allele GG displayed a red hue. Similar results were noted in the other species, where the color variation in ripe pepper fruit extended from orange to red. Additionally, we used SNP pairs from GWAS to develop KASP assays consisting of two allele-specific primers and a common reverse primer for selected markers 01:31205460 and 08:45351919 located on PPR-containing protein and Cytochrome P450, respectively (Table S9). The markers resolved three distinct groups: homozygous allele 1, heterozygous, and homozygous allele 2, which showed that the marker genotype and phenotype were utterly consistent with color traits (Fig. 7). These results indicate that markers 01:31205460 and 08:197099127 can be applied in breeding programs for high-quality colored pepper varieties in the future. However, these markers should be further validated in all *Capsicum* species.

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).

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 spaces 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 genes encoding key enzymes and their transporters make them critical components in the broader landscape of fruit coloration and phytonutrient content in pepper fruits.

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. 2021; 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 either of 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. 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. PPRs are involved in RNA modifications, including RNA editing, RNA stabilization, and splicing (Qin et al. 2021), suggesting a possible role of RNA modifications in regulating fruit color traits in pepper.

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 embarks on a transformative journey into the genetic intricacies of color variation within four cultivated species of the *Capsicum* genus. This research delineates causal SNPs through a precise and systematic approach. It sheds 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 tapestry of hues displayed in *Capsicum* fruits, 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. By targeting specific color attributes, breeders can now harness these markers to sculpt the chromatic palette of pepper fruits with unprecedented precision. Anticipating future research avenues, an expansive GWAS that scrutinizes secondary metabolites, mainly carotenoids and anthocyanins, is proposed. These metabolites are the maestros of color in chili peppers, orchestrating a visual symphony that is as appealing to consumers as it is to researchers. The insights gleaned from this study cast light on the genetic architecture that determines the vibrant spectrum of *Capsicum* fruit colors. The gene families pinpointed here play significant roles in determining the colors of crops such as apples, grapes, rice, and maize. This study also provides a genetic repository that can be tapped into for enhancing the aesthetic, nutritional, and commercial value of pepper varieties and other agriculturally important crops, highlighting the potential to influence global food systems.

Declarations

Author Contributions

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

Conflicts of Interest

The authors declare no conflict of interest.

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).

Acknowledgments

The authors thank Venkata Abburi, Krittika Tonapi, and Sai Satish Kadiyala for field studies, technical assistance, and help with the analysis and compiling.

References

1. Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC (2009) Genetic diversity and structure in semiwild and domesticated chiles (Capsicum annuum; Solanaceae) from Mexico. American Journal of Botany 96:1190-1202

2. 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 Journal 28:319-332
3. 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 10
4. 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 Physiology 178:808-823
5. Azlan A, Sultana S, Huei CS, Razman MR (2022) Antioxidant, Anti-Obesity, Nutritional and Other Beneficial Effects of Different Chili Pepper: A Review. Molecules 27
6. 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 x Solanum chmielewskii Introgression Line Population. Frontiers in Plant Science 7
7. 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
8. Borges RM (2015) Fruit and seed volatiles: multiple stage settings, actors and props in an evolutionary play. Journal of the Indian Institute of Science 95:93-104
9. 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. Theoretical and Applied Genetics 109:23-29
10. 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. Theoretical and Applied Genetics 117:235-240
11. 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). Annals of botany 118:35-51
12. 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 Physiology 161:225-239
13. 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. Horticulture Advances 1:5
14. 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 Physiology and Biochemistry 162:267-279
15. 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 7

16. 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
17. 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
18. 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.). Theoretical and Applied Genetics 134:3351-3362
19. Craig KL, Tyers M (1999) The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Progress in Biophysics & Molecular Biology 72:299-328
20. 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. Journal of the American Society for Horticultural Science 133:579-586
21. 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 Science and Technology 40:121-129
22. 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. Scientific Reports 13
23. 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. Frontiers in Plant Science 6:1233
24. 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
25. 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 Physiology 136:4184-4197
26. 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 Physiology 169:1714-1726
27. 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.). Molecular Biology Reports 49:5283-5291
28. 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. Frontiers in Plant Science 12
29. 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

30. 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
31. 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. The Plant Journal 94:169-191
32. Gonzali S, Perata P (2021) Fruit Colour and Novel Mechanisms of Genetic Regulation of Pigment Production in Tomato Fruits. Horticulturae 7
33. 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 Physiology 176:2292-2304
34. He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. Crop breeding: methods and protocols:75-86
35. 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
36. 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. Journal of Agricultural and Food Chemistry 48:1713-1720
37. Hurtado-Hernandez H, Smith PG (1985) INHERITANCE OF MATURE FRUIT COLOR IN CAPSICUM-ANNUUM-L. Journal of Heredity 76:211-213
38. Jin HL, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Molecular Biology 41:577-585
39. Kapoor L, Simkin AJ, Doss CGP, Siva R (2022a) Fruit ripening: dynamics and integrated analysis of carotenoids and anthocyanins. Bmc Plant Biology 22
40. Kapoor L, Simkin AJ, George Priya Doss C, Siva R (2022b) Fruit ripening: dynamics and integrated analysis of carotenoids and anthocyanins. BMC plant biology 22:1-22
41. 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 physiologiae plantarum 35:2879-2890
42. Kiferle C, Fantini E, Bassolino L, Povero G, Spelt C, Buti S, Giuliano G, Quattrocchio F, Koes R, Perata P, Gonzali S (2015) Tomato R2R3-MYB Proteins SlANT1 and SlAN2: Same Protein Activity, Different Roles. Plos One 10
43. 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 Biology 18

44. 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. Nature Genetics 46:270-+
45. 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. Proceedings of the National Academy of Sciences 111:6165-6170
46. 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 Biology 10
47. 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
48. 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. Frontiers in Genetics 11
49. 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 Journal
50. Mendes ND, Goncalves E (2020) The role of bioactive components found in peppers. Trends in Food Science & Technology 99:229-243
51. 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 Physiology and Biochemistry 89:24-30
52. Moreno JC, Mi JN, Alagoz Y, Al-Babili S (2021) Plant apocarotenoids: from retrograde signaling to interspecific communication. Plant Journal 105:351-375
53. Munshell AH (1912) A pigment color system and notation. American Journal of Psychology 23:236-244
54. 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 Bioscience:102403
55. 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 9

56. 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. Scientia Horticulturae 260
57. 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
58. 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. Frontiers in Genetics 14
59. Philippe F, Pelloux J, Rayon C (2017) Plant pectin acetylesterase structure and function: new insights from bioinformatic analysis. Bmc Genomics 18
60. Pickersgill B (1971) Relationships between weedy and cultivated forms in some species of chili peppers (genus Capsicum). Evolution 25:683-691
61. 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. Proceedings of the National Academy of Sciences of the United States of America 111:5135-5140
62. 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. Frontiers in Genetics 12
63. 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
64. 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. Journal of Experimental Botany 63:517-526
65. Sathasivam R, Ki J-S (2018) A review of the biological activities of microalgal carotenoids and their potential use in healthcare and cosmetic industries. Marine drugs 16:26
66. Schifferstein HNJ, Wehrle T, Carbon CC (2019) Consumer expectations for vegetables with typical and atypical colors: The case of carrots. Food Quality and Preference 72:98-108
67. 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. Theoretical and Applied Genetics 124:295-307

68. 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
69. 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. Critical Reviews in Plant Sciences 38:339-381
70. 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. International Journal of Molecular Sciences 22
71. Shumskaya M, Wurtzel ET (2013) The carotenoid biosynthetic pathway: Thinking in all dimensions. Plant Science 208:58-63
72. 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
73. 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 11
74. Stommel JR, Dumm JM (2015) Coordinated regulation of biosynthetic and regulatory genes coincides with anthocyanin accumulation in developing eggplant fruit. Journal of the American Society for Horticultural Science 140:129-135
75. Stommel JR, Lightbourn GJ, Winkel BS, Griesbach RJ (2009) Transcription Factor Families Regulate the Anthocyanin Biosynthetic Pathway in Capsicum annuum. Journal of the American Society for Horticultural Science 134:244-251
76. 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 11
77. 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 Physiology 142:1216-1232
78. Tanaka Y, Brugliera F (2013) Flower colour and cytochromes P450. Philosophical Transactions of the Royal Society B-Biological Sciences 368
79. 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. Horticulture Research 9:uhac072
80. Villa-Rivera MG, Ochoa-Alejo N (2021) Transcriptional Regulation of Ripening in Chili Pepper Fruits (Capsicum spp.). International Journal of Molecular Sciences 22

81. Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG (2013) Secondary Metabolites of Capsicum Species and Their Importance in the Human Diet. Journal of Natural Products 76:783-793
82. 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. International Journal of Plant Sciences 162:1409-1418
83. 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) 12
84. Wang XW, An YQ, Xu P, Xiao JW (2021) Functioning of PPR Proteins in Organelle RNA Metabolism and Chloroplast Biogenesis. Frontiers in Plant Science 12
85. Weale ME (2010) Quality Control for Genome-Wide Association Studies. In: Barnes MR, Breen G (eds) Genetic Variation: Methods and Protocols, pp 341-372
86. Williams PM, Barkan A (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. Plant Journal 36:675-686
87. 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 14
88. 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. Scientia Horticulturae 323:112502
89. Xu J, Wang XY, Guo WZ (2015) The cytochrome P450 superfamily: Key players in plant development and defense. Journal of Integrative Agriculture 14:1673-1686
90. 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 Science 23:1-8
91. 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. Frontiers in Plant Science 13

Figures

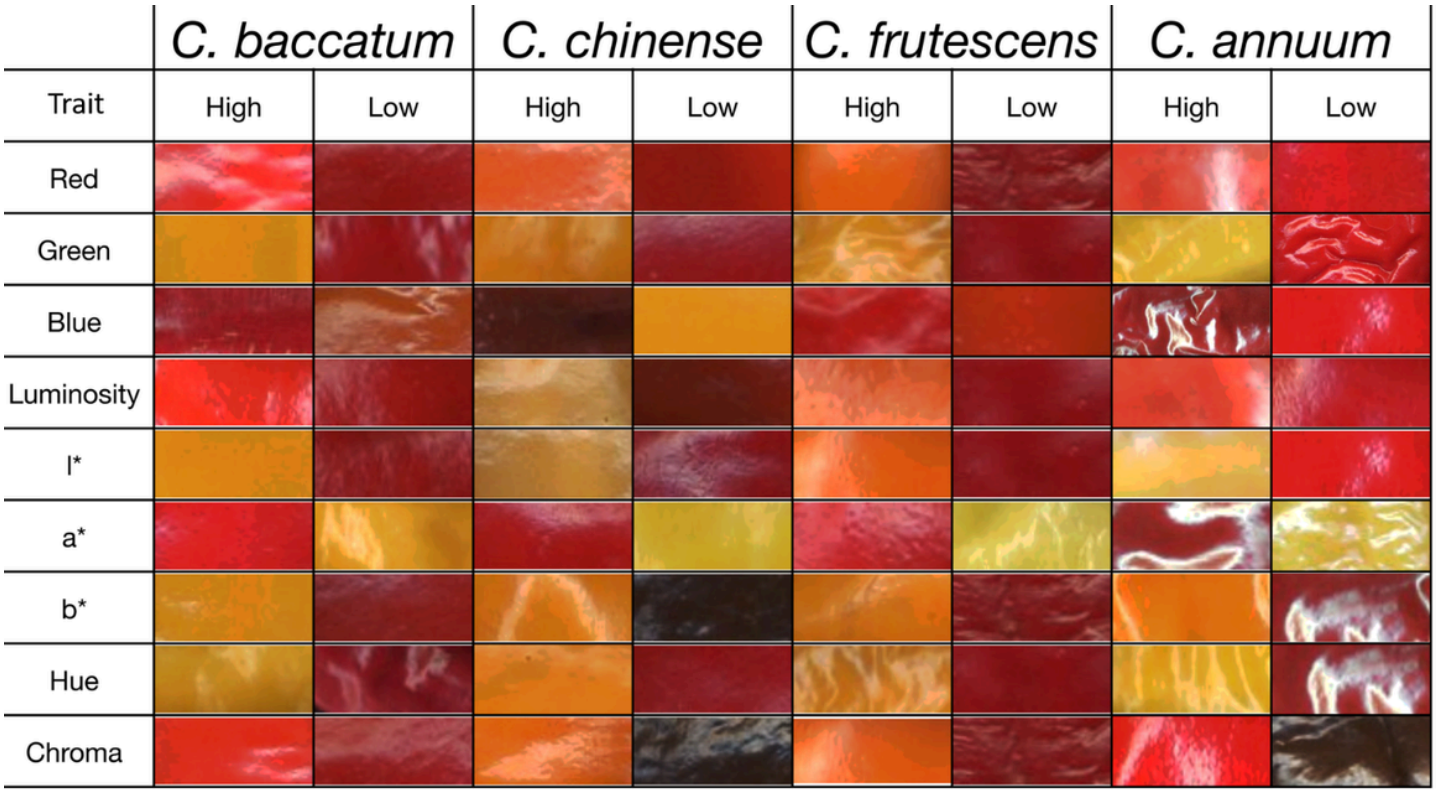

Figure 1

Phenotypic diversity of 9 color spaces across *Capsicum* Species. The color spaces provide a comprehensive view of the rich spectrum of colors found in *Capsicum*peppers, highlighting the intriguing variations in fruit pigmentation among these pepper species.

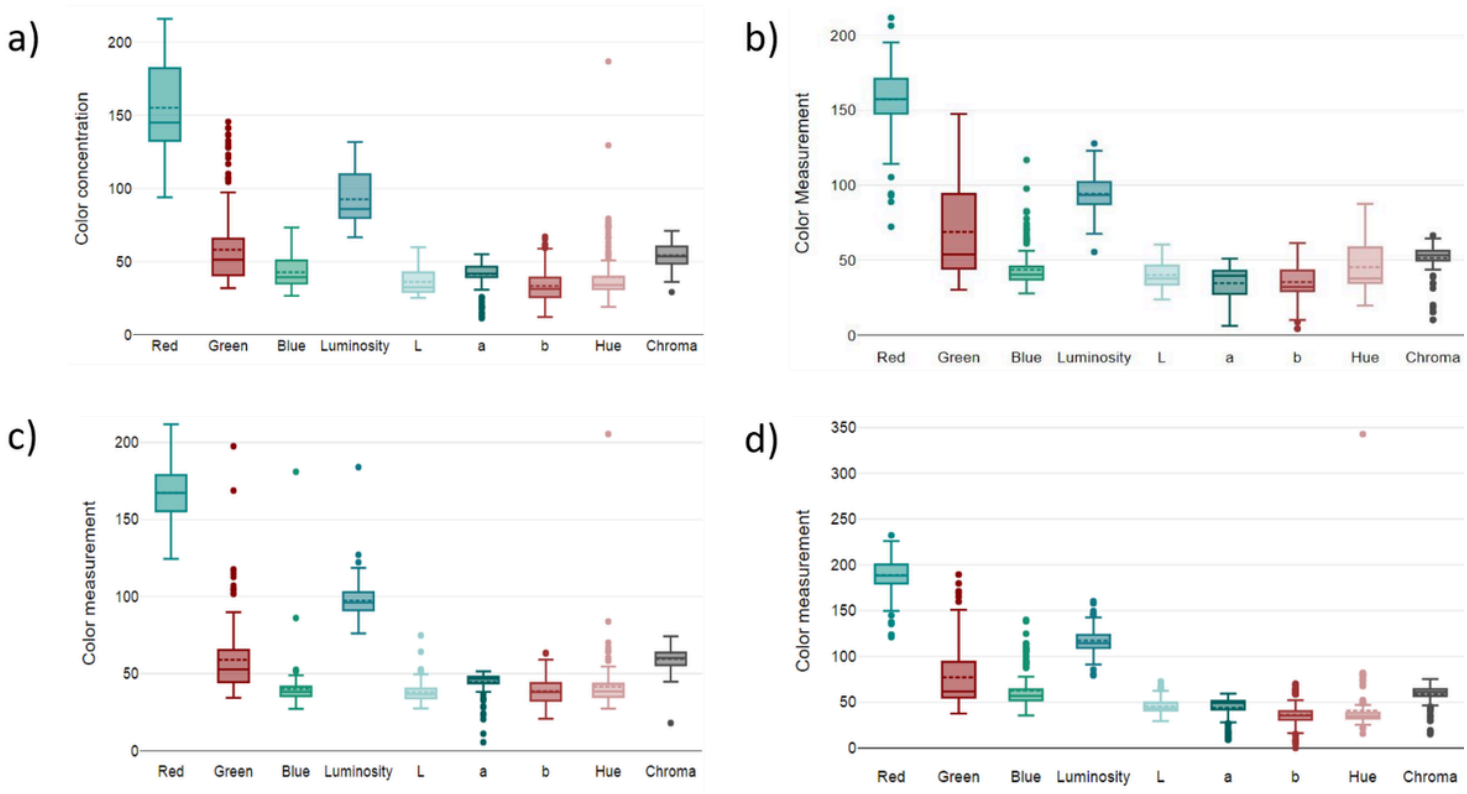

Figure 2

Box plots for nine color fruit agronomic traits of a) *C. baccatum*, b) *C. chinense,* c) *C. frutescens,* and d) *C. annuum* accessions grown in field conditions during 2015-2021 seasons.

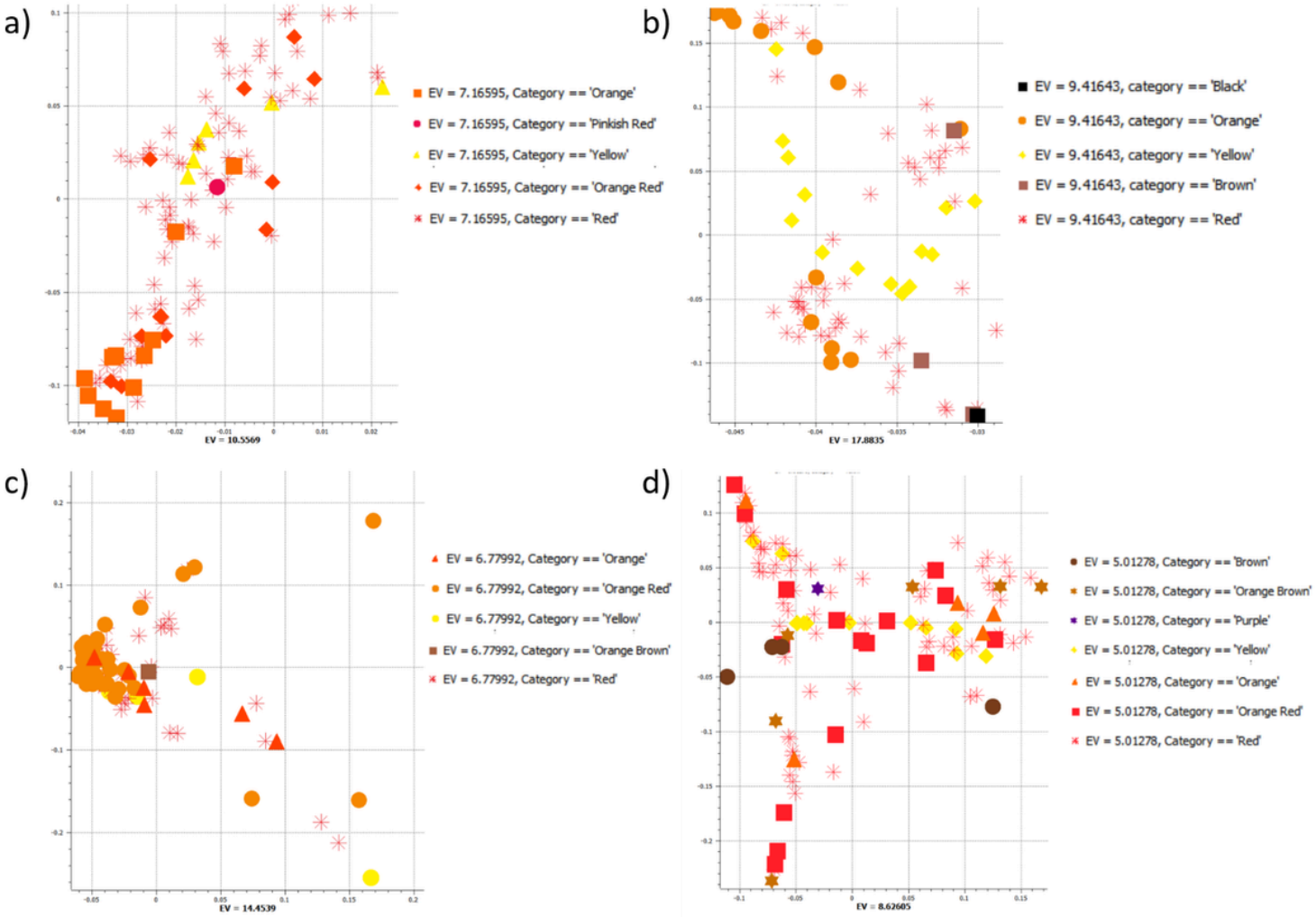

Figure 3

Principal component analysis showing the genetic distribution of a) *C. baccatum*, b) *C. chinense,* c) *C. frutescens,* and d) *C. annuum* accessions collections according to single nucleotide polymorphisms (SNPs).

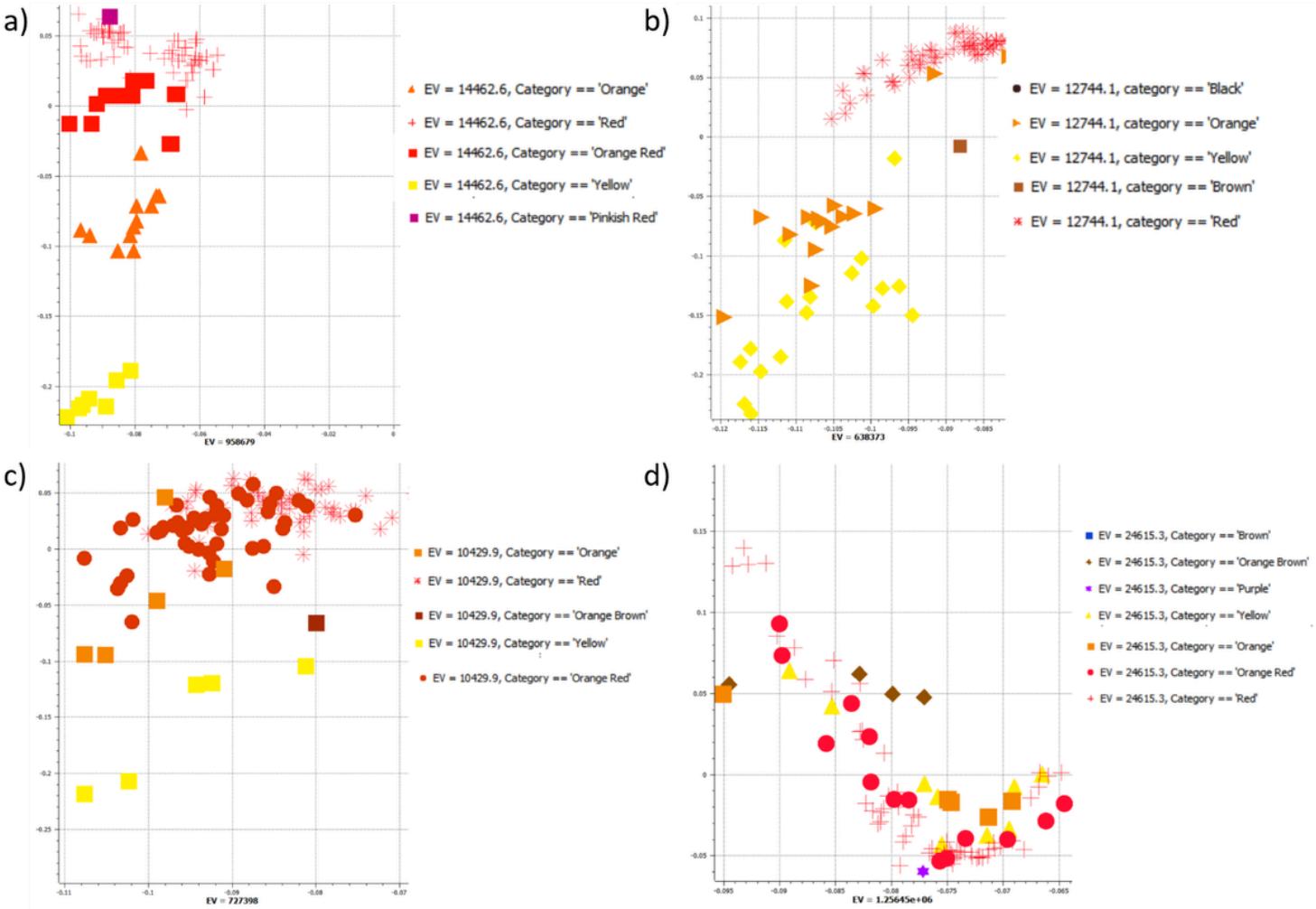

Figure 4

Principal component analysis showing the genetic distribution of a) *C. baccatum*, b) *C. chinense,* c) *C. frutescens,* and d) *C. annuum* accessions collections according to numerical color values.

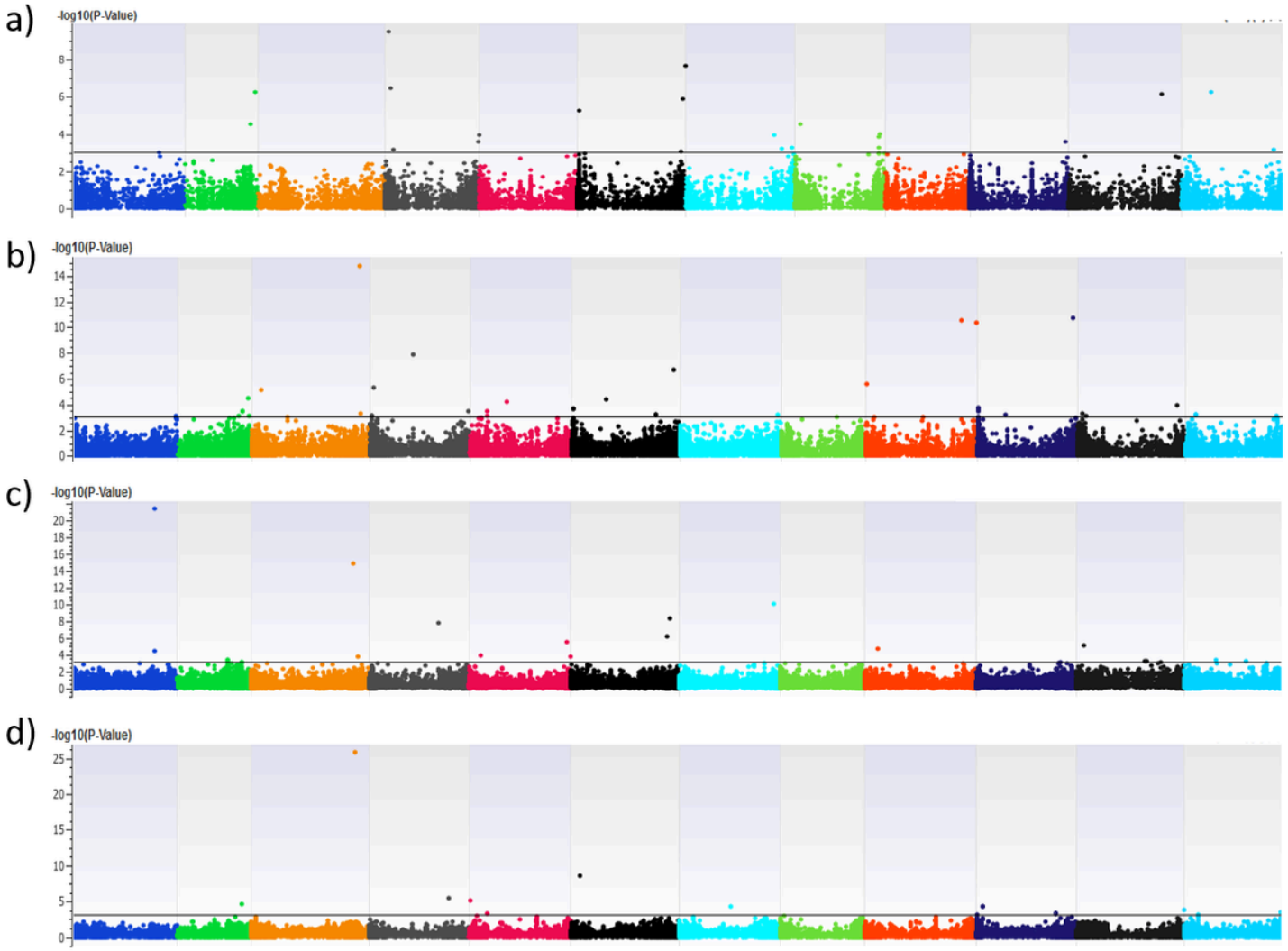

Figure 5

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. Black line indicates a threshold of −log10 (*P*-value) = 3 at *P* = 0.05 after false discovery rate correction. Red arrows show highly significant associated SNPs.

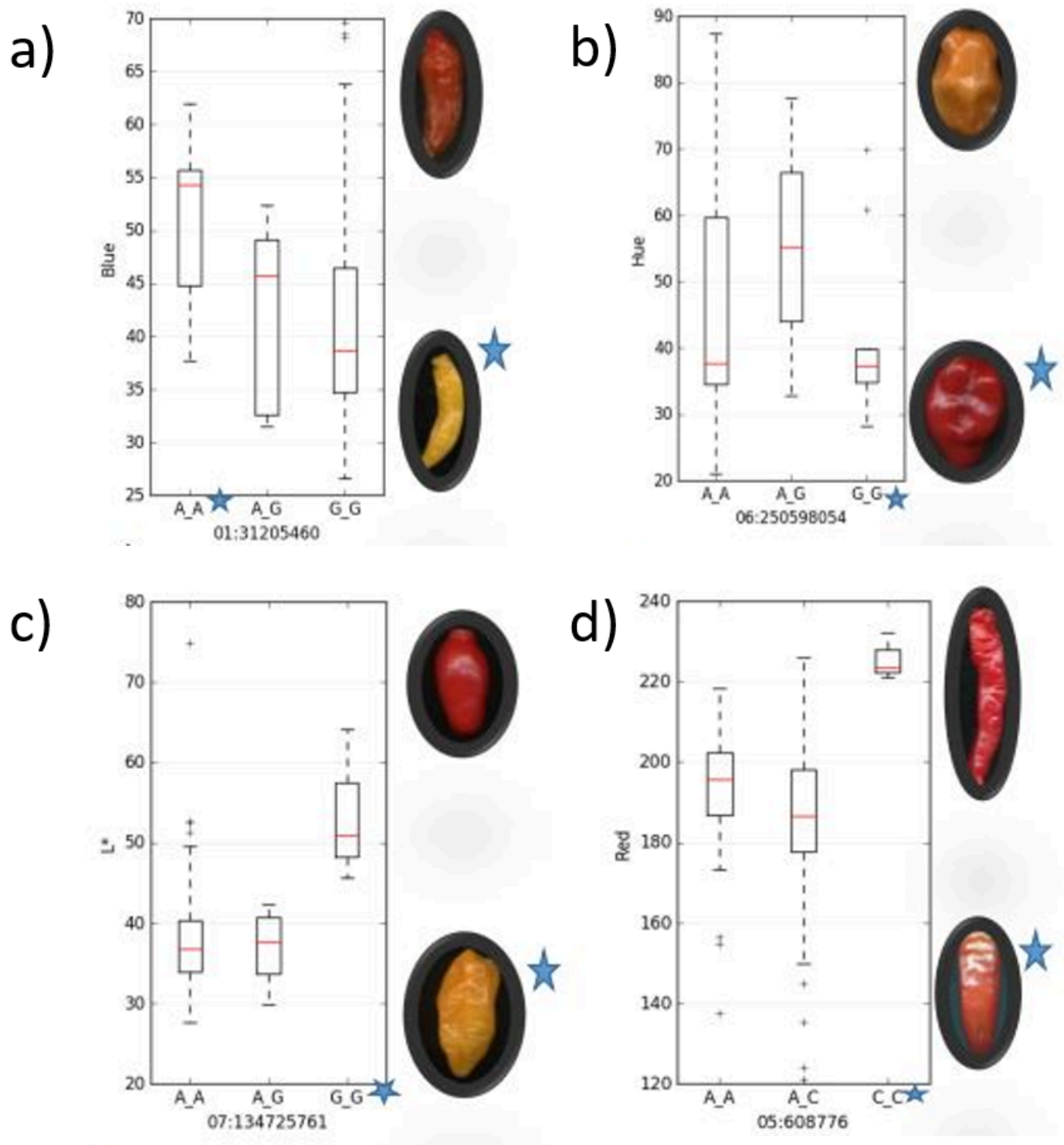

Figure 6

Allelic effect of significantly associated SNP markers for Pentatricopeptide repeat-containing protein in a) *C. baccatum* on chromosome 01, b) *C. chinense* on chromosome 06, c) *C. frutescens* on chromosome 07, and d) *C. annuum* on chromosome 05. Box plots show the effect of the SNP marker, and the Y-axis represents values for each color trait. The star symbol represents the major allele.

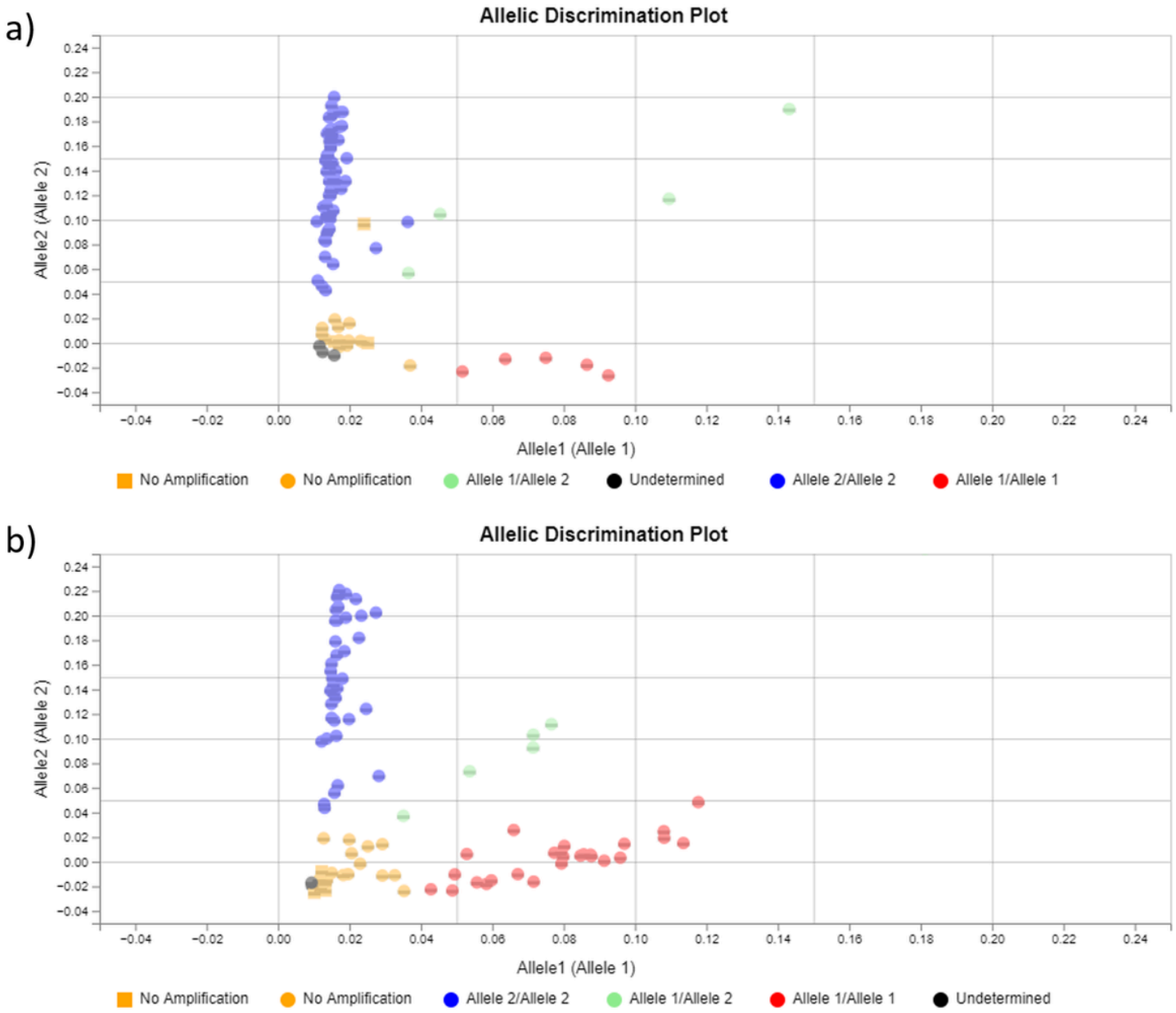

Figure 7

Genotypes of selected KASP markers a) Chr01:31205460 (pentatricopeptide repeat-containing protein) and b) 08:45351919 (cytochrome P450) in the *C. baccatum* population. 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). The red and blue fluorescence intensity was determined in a StepOne Plus system.

Supplementary Files

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

- SupplementaryFigures.docx

- SupplementaryTables.xlsx