Anthocyanin gene enrichment in the distal region of cotton chromosome A07: mechanisms of reproductive organ coloration

Liuchang Zheng[1,2†], Jilong Zhang[2†], Haiyan He[2†],
Zhigang Meng[2†], Yuan Wang[2], Sandui Guo[2]
and Chengzhen Liang[2*]

[1]College of Life Sciences, Qingdao Agricultural University, Qingdao, China, [2]Biotechnology Research
Institute, Chinese Academy of Agricultural Sciences, Beijing, China

Introduction: The biosynthesis of secondary metabolites like anthocyanins is often governed by metabolic gene clusters (MGCs) in the plant ancestral genome. However, the existence of gene clusters specifically regulating anthocyanin accumulation in certain organs is not well understood.

Methods and results: In this study, we identify MGCs linked to the coloration of cotton reproductive organs, such as petals, spots, and fibers. Through genetic analysis and map-based cloning, we pinpointed key genes on chromosome A07, such as *PCC/GhTT19*, which is involved in anthocyanin transport, and *GbBM* and *GhTT2-3A*, which are associated with the regulation of anthocyanin and proanthocyanidin biosynthesis. Our results demonstrate the coordinated control of anthocyanin and proanthocyanidin pathways, highlighting the evolutionary significance of MGCs in plant adaptation. The conservation of these clusters in cotton chromosome A07 across species underscores their importance in reproductive development and color variation. Our study sheds light on the complex biosynthesis and transport mechanisms for plant pigments, emphasizing the role of transcription factors and transport proteins in pigment accumulation.

Discussion: This research offers insights into the genetic basis of color variation in cotton reproductive organs and the potential of MGCs to enhance our comprehension of plant secondary metabolism.

KEYWORDS

cotton, anthocyanin, MYB, GST, metabolic gene clusters

Introduction

In plants, the genomic architecture for the biosynthesis of secondary metabolites is characterized by the organization of relevant genes into metabolic gene clusters (MGCs). These clusters facilitate the synthesis of critical compounds that enhance plant resilience against diverse biotic stresses (Jonczyk et al., 2008; Chu et al., 2011; Sue et al., 2011; Takos et al., 2011; Guo et al., 2017).

Plant MGCs can be classified into two distinct configurations. The first configuration dominantly is tightly packed gene clusters, wherein genes responsible for both the initial and subsequent enzymatic reactions are sequentially arranged. This arrangement is exemplified by the thalianol gene cluster in *Arabidopsis* and the diterpenoid gene cluster in rice, illustrating a streamlined genomic strategy for secondary metabolite biosynthesis (Liu et al., 2020; Zhan et al., 2020). The second configuration involves partially adjacent and dispersed gene clusters, where a core group of genes is closely located, albeit with several genes situated distantly. The steroidal alkaloid biosynthesis in tomatoes follows the α-tomatine pathway, controlled by a gene cluster on chromosome 7. This cluster includes two 2-oxoglutaratedependent dioxygenase (*GAME11* and *GAME6*)genes, four genes encoding glycosyltransferase (*GAME1,GAME17,GAME18* and *GAME2*), and the cytochrome P450 monooxygenase gene (*GAME7*), separated by 8 Mb from the nearest gene cluster (Itkin et al., 2013). Similarly, cucumber's cucurbitacin biosynthesis is managed by a gene cluster on chromosome 6, with additional cytochrome P450 (CYP) genes on chromosomes 3 and 1, indicating a spread but functionally unified genomic organization (Shang et al., 2014). In potatoes, sterol biosynthesis genes are found on chromosomes 7 and 12, highlighting the varied arrangement of MGCs in different plant species (Cardenas et al., 2016).

The biosynthesis pathways of anthocyanins and proanthocyanidins may also be regulated by the presence of MGCs, underscore the evolutionary significance of such genomic configurations for plant adaptation and survival. Anthocyanins, alongside carotenoids and betalains, are pivotal for the vibrant pigmentation observed in plant tissues, with anthocyanins occupying a major role in coloration (Tanaka et al., 2008). These pigments are predominantly accumulated in essential reproductive structures—flowers, fruits, and seeds—creating diverse, vivid color patterns that both attract pollinators and confer resistance to biotic and abiotic stressors (Stintzing and Carle, 2004; Duan et al., 2014). Anthocyanin biosynthesis is mediated by a sequence of enzymatic reactions on the endoplasmic reticulum (ER), initiating from phenylalanine (Winkel, 2004; Yang et al., 2018; Belwal et al., 2020). Meanwhile, this pathway further promoted the synthesis of proanthocyanidins (PAs), which affected the pigmentation across various plant organs (Shi et al., 2021; Zhao et al., 2023).

Genes involved in the biosynthesis of anthocyanins and proanthocyanidins have been identified across various plant species (Winkel-Shirley, 2001; Ma et al., 2021; Luan et al., 2022; Yang et al., 2022; He et al., 2023). The anthocyanin biosynthetic pathway encompasses three main steps: Initially, phenylalanine is converted into 4-Coumaroyl-CoA via the phenylalanine ammonia-lyase (PAL), cinnamic acid hydroxylase (C4H), and 4-coumarate-CoA ligase (4CL) enzymes. Subsequently, 4-Coumaroyl-CoA is transformed into dihydroflavonol, a shared precursor for both anthocyanins and proanthocyanidins, through the action of chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3-hydroxylase (F3H). Dihydroflavonol is then converted into the colorless leucoanthocyanidins (leucopelargonidin, leucocyanidin, and leucodelphinidin) by dihydroflavonol reductase (DFR). In the final synthesis phase, anthocyanins are produced from these colorless precursors via anthocyanin synthase (ANS) and leucoanthocyanidin dioxygenase (LODX), with subsequent glycosylation and acylation by UDP-3-O-glucosyltransferase (UFGT) to yield stable anthocyanins. Alternatively, colored anthocyanins can be reduced to epicatechins by anthocyanin reductase (ANR), or directly to catechins by leucoanthocyanidin reductase (LAR), which then polymerize to form proanthocyanidins in the vacuole (Zhao et al., 2010; Yu et al., 2023). The activity of these enzymes is regulated by v-myb avian myeloblastosis viral oncogene homolog (MYB), WD40, and basic Helix-Loop-Helix (bHLH) transcription factors, highlighting a complex control mechanism over the biosynthetic pathway (Petroni and Tonelli, 2011; Yan et al., 2021; Shi et al., 2023).

The transport of anthocyanins, catechins, and proanthocyanidins to vacuoles can occur through two pathways: vesicle transport facilitated by vesicle wrapping and transport proteins, including ATP-binding cassette transporters (ABC/MRP), multidrug and toxic compound extrusion (MATE), multidrug resistance-associated protein (MRP), and glutathione S-transferase (GST) (Grotewold and Davies, 2008; Gomez et al., 2011). GST proteins play a crucial role in the transport of anthocyanins and monomers of anthocyanins. Acting as transport carriers, GST proteins directly bind to anthocyanins or precursors of proanthocyanidins in the cytoplasm, facilitating their transport to the vacuolar membrane. Subsequently, these compounds are translocated across the membrane into the vacuole after recognition by ABC/MRP or MATE transmembrane proteins located on the vacuolar membrane (Loyall et al., 2000; Zhao, 2015).

The genus *Gossypium*, encompassing over fifty species, exhibits significant diversity in flower and fiber coloration, primarily attributed to differential gene expression within the anthocyanin/proanthocyanidin biosynthetic pathways (Tan et al., 2013). Notably, key genes implicated in the regulation of these pathways, such as *G. barbadense Beauty Mark* (*GbBM*), *G. hirsutum TRANSPARENT TESTA 2–3A* (*GhTT2-3A*) in reproductive tissues, are members of the MYB transcription factor family (Yan et al., 2018; Abid et al., 2022). Additionally, the glutathione S-transferase gene *GhTT19* plays a pivotal role in the transport of anthocyanins/proanthocyanidins, influencing flower petal pigmentation (Chai et al., 2023). Intriguingly, these genes associated with the synthesis and transports of anthocyanins/proanthocyanidins in cotton reproductive organs are clustered in the distal region of chromosome A07 in cotton. The enrichment of these genes in specific chromosomal regions related to anthocyanin synthesis and transport remains largely unexplored. Our study demonstrates a significant correlation between the distal region of chromosome A07 and the pigmentation of cotton reproductive organs, suggesting the formation of a metabolic gene cluster that regulates the anthocyanin pathway, thereby influencing coloration. This chromosomal configuration remains conserved across the transition from diploid to tetraploid species, underscoring its importance in cotton reproductive development and propagation. These insights offer a foundation for future research aimed at enhancing cotton hybrid breeding, the cultivation of naturally colored fibers, and improving cotton yield.

Materials and methods

Plant materials

The *G. barbadense* line HaiR, *G. hirsutum* line Y18R, and red color cotton line *pcc*, were used in this study. The red-petaled *pcc* cotton was selected from over 1,000 varieties in the cotton germplasm resource library. The Y18R line is characterized by white flowers, whereas the *pcc* line exhibits red flowers. The HaiR line is distinguished by its yellow flowers. For the generation of F₂ mapping populations, HaiR and *pcc* were selected as the parental lines.

Cotton planting conditions

Cotton plants were grown under field conditions at the High-Tech Industry Park of the Chinese Academy of Agricultural Sciences in Langfang, Hebei Province (N39°520, E116°700E). Photographic documentation and petal collection were conducted during and post-flowering for subsequent analyses. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) and anthocyanin content determination were performed, with each assay conducted in triplicate.

Map-based cloning

Map-based cloning was conducted using an F₂ population consisting of a total of 870 cotton plants, which exhibited two distinct phenotypes: yellow-flowered and red-flowered. For genetic linkage analysis, 270 individuals displaying the yellow color flower phenotype were selected at the flowering stage. A total of 720 pairs of insertion/deletion (InDel) primers, designed based on genome-wide InDel markers differentiating *G.barbadense* and *G.hirsutum* genotypes, were used for preliminary gene mapping (He et al., 2018). The PCR amplifications were electrophoresed on a 4% agarose gel. Primers utilized for fine mapping are detailed in Supplementary Table 3.

Draw the anthocyanin metabolism pathway

Building on previous research into the metabolic pathways of anthocyanins and proanthocyanidins (Liu et al., 2021; Pucker and Selmar, 2022; Sunil and Shetty, 2022), and we made minor adjustments and illustrated these pathways using Figdraw to accurately depict the metabolic regulation model.

RNA extraction and qRT-PCR analysis

For expression analysis, samples were collected and immediately flash-frozen in liquid nitrogen, then stored at -80°C until RNA extraction. Total RNA was extracted using the RNAprep Pure Plant Kit (TIANGEN, DP441, Beijing, China) following the manufacturer's protocol. cDNA synthesis was performed using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TRAN, AT311, Beijing, China) as per the manufacturer's guidelines. qRT-PCR analysis was conducted using the Chromo 4 Real-Time PCR Detection System (Bio-Rad, CFX96, Hercules, USA), adhering to the manufacturer's instructions. The cotton *GhHistone* gene served as the internal reference. Experiments were independently replicated three times. Primers for qRT-PCR are detailed in Supplementary Table 3. The relative expression levels of genes were quantified using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Gene expression analysis

The raw RNA-seq datasets for red and white flowers of *G. hirsutum* were obtained from PRJNA878950, while those for brown and white cotton fibers were sourced from PRJNA766762. Following the removal of low-quality reads, the remaining clean reads were aligned to the TM-1 reference genome using Hisat2 software (Kim et al., 2019). Differential gene expression analysis was conducted using featureCounts for the calculation of Transcripts Per Million (TPM) values (Liao et al., 2014; Zhao et al., 2020). Gene expression heatmaps were generated utilizing the Pheatmap package (Hu, 2021).

Anthocyanin analysis

The determination of anthocyanin content followed the methodology described by Xu et al. (2012). Petal samples (0.1 g) were retrieved from a -80°C freezer, ground to a powder in liquid nitrogen, and then mixed with 1 mL of methanol. The mixture was vortexed thoroughly and sonicated for 15 minutes. After centrifugation at 12,000 rpm for 10 minutes, the supernatant was diluted tenfold with methanol and filtered through a 0.22 μm organic membrane. LC-MS analysis was performed using a Thermo U3000 HPLC System. The peak area of the mixture was quantified by comparison with the peak areas of standard substances, including cyanidin, delphinidin, and pelargonidin (all from Sigma, USA). The formula for calculating anthocyanin content is as follows: [A = (C2 - C1) * V * N where (A) represents the content of various anthocyanins in the sample (mg/kg), (C2) is the concentration of each anthocyanin in the test solution (mg/L), (C1) is the concentration of each anthocyanin in the blank solution (mg/L), (V) is the volume of the solution (mL), and (N) is the dilution factor.

Population structure analysis

Total 96 accessions derived from four projects (PRJNA349094, PRJNA680449, PRJNA412456, PRJNA720818) were analyzed (see Supplementary Table 4). The sequencing reads were aligned to the *G. hirsutum* TM-1 reference genome using BWA software (Li and

Durbin, 2009). The resulting BAM files were processed for sorting and filtering using SAMtools (version 1.15) with default parameters (Danecek et al., 2021). SNP calling at the population level was conducted using GATK software (version 4.0) (Van der Auwera et al., 2013), generating a GVCF file for each sample. These GVCF files were then merged, applying the following parameters for filtering: –filter-expression "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" – filter-name 'SNP_filter'.

A total of 120,126 SNPs, with a minor allele frequency (MAF) > 0.01 and missing data < 10%, were identified using Plink (version 1.9). These SNPs were used to construct a phylogenetic tree using the neighbor-joining method in PHYLIP software (version 3.698) (Retief, 2000). The phylogenetic tree was visualized and refined using iToL (https://itol.embl.de/). Population genetic structure was assessed using Admixture (Version 1.3.0), with the number of genetic clusters (K) ranging from 2 to 6 and each run consisting of 10,000 iterations.

Analysis of the domestication selection

Using VCFtools (version 0.1.16) (Danecek et al., 2011), we calculated fixation index (Fst) and nucleotide diversity (π) employing a sliding window approach. Each window spanned 50kb, with successive windows overlapping by 10kb. This methodology facilitates a detailed and nuanced analysis of genetic differentiation and diversity across the genome.

Identification of MYB and GST gene families

Genomic sequences, annotation files, and protein sequences for *Gossypium* species including *G. herbaceum* (A1), *G. arboretum* (A2), *G. raimondii* (D5), *G. hirsutum* (AD1), *G. barbadense* (AD2), *G. tomentosum* (AD3), *G. mustelinum* (AD4), and *G. darwinii* (AD5) were retrieved from Phytozome (Goodstein et al., 2012) and CottonMD (Yang et al., 2023). Protein sequences for the GST and MYB gene families were downloaded from the TAIR database (https://www.arabidopsis.org/). Hidden Markov Model (HMM) profiles for GST_C (PF00043), GST_N (PF02798), and MYB (PF00249) were acquired from the Pfam database (version 35.0).

To identify homologs of the MYB and GST gene families across the *Gossypium* species, *Arabidopsis thaliana* GST and MYB protein sequences served as query sequences in BLASTp searches (version 2.15), with stringent criteria set at e-values < 1e-10 and identity ≥ 30%. Concurrently, HMMER (version 3.4) software, utilizing the hmmersearch function and the aforementioned HMM profiles, was employed to search for protein sequences within the cotton varieties, applying an e-value threshold of ≤ 1e-5. Results from both BLASTp and hmmersearch were integrated, allowing for the exclusion of structurally incomplete sequences from the GST and MYB gene families in the analyzed *Gossypium* species.

Multiple sequence alignment and phylogenetic tree construction

Protein sequences from various *Gossypium* species were consolidated and subjected to multiple sequence alignment using MUSCLE (version 5.0). Subsequently, a multi-species evolutionary tree was constructed employing the Maximum Likelihood (ML) method via FastTree (Price et al., 2009). The resulting evolutionary tree was visualized and refined using the online tool iTOL (Letunic and Bork, 2021). Within this phylogenetic framework, we specifically identified protein sequences associated with anthocyanin biosynthesis. These sequences were then realigned using MUSCLE (version 5.0), and the alignment was further refined for accuracy and clarity using GeneDoc.

Chromosomal localization, duplication, and synteny analysis

Protein sequences from all varieties were compiled into a database, with each serving as its own query sequence. For sequence alignment, Diamond (version 2.1.8) (Buchfink et al., 2021) was utilized, applying an e-value threshold of < 1e-10. The gene annotation files were then converted into GFF format. Duplication classification of genes was conducted using McscanX (Wang et al., 2012), which produced source files detailing gene family duplications. A Perl script, get_mapchart_tandem.pl, was used to generate configuration files for MapChart (Voorrips, 2002), which depicted the genomic locations and tandem duplications of the GST and MYB gene families on chromosomes.

For inter-species collinearity analysis, MCScanX was used to create collinearity files. These files were then modified using another Perl script, highlight.pl, to emphasize blocks containing GST and MYB genes. The visualization of these results was achieved through the JCVI software (Tang et al., 2008), effectively mapping the positions and tandem duplications of the GST and MYB gene families, as well as highlighting significant collinearity blocks.

Construct protein regulatory network

To elucidate the relationship between MYB genes on Chromosome A07 and structural genes involved in anthocyanin and proanthocyanidin biosynthesis, we employed the String database (https://string-db.org/) for protein-protein interaction (PPI) prediction. The resultant PPI predictions were then visualized through the construction of a network diagram using Cytoscape software (version 3.10.1) (Otasek et al., 2019).

Data analysis

To compare two groups, the Student's t-test was utilized. For analyzing significance across multiple groups, one-way ANOVA followed by Tukey's multiple comparisons test was employed. Data

Results

The distal region of chromosome A07 in cotton regulates genes for reproductive organ coloration

The synthesis of plant anthocyanins is significantly influenced by transcriptional regulatory factors, primarily MYB, bHLH, and WD40, which together form a regulatory system that controls anthocyanin biosynthesis by binding to structural gene promoters. Once synthesized on the endoplasmic reticulum surface, anthocyanins are transported into the vacuole for storage, facilitated by Glutathione S-transferase (GST) and recognized by multidrug resistance associated protein transporter and ABC protein transporter on the vacuolar membrane (Figure 1A). We previously reported a gene, *GbBM*, which encodes an R2R3 MYB113 transcription factor. This factor directly targets the promoters of four flavonoid biosynthesis genes (*GbCHS*, *GbDFR*, *GbANS*, and *GbUFGT*), thereby positively regulating the formation of purple spots on cotton flower petals (Abid et al., 2022). In recent years, leveraging map-based cloning techniques, we identified the gene *PCC* from a red-flowered cotton resource, which governs petal coloration in cotton (Supplementary Figure 1 and Supplementary Table 1). *PCC*, previously known as *GhTT19*, encodes a typical glutathione S-transferase. This enzyme is pivotal in petal pigmentation, as it aids in the transport and accumulation of anthocyanins. Significantly, *GbBM* and *PCC/GhTT19* not only share the function of regulating petal coloration but are also positioned closely at the proximal end of the cotton A07 chromosome, separated by a minimal physical distance of merely 0.85MB (Supplementary Table 2). This proximity underscores a potentially significant genetic linkage in the control of floral pigmentation.

To investigate the potential association between the proximal region of the cotton A07 chromosome and pigment development, we examined genes responsible for the coloration of reproductive tissues. Notably, our findings revealed that *GhTT2-3A*, a gene implicated in the production of brown fibers through the accumulation of anthocyanins/proanthocyanidins, is also situated in this specific chromosomal region. *GhTT2-3A*, a homolog of the *Arabidopsis* PA regulator TT2 within the R2R3 MYB TF family, triggers the activation of downstream PA structural genes (Yan et al., 2018). This activation facilitates the synthesis and accumulation of proanthocyanidins (PAs) in cotton fibers, leading to the emergence of brown-colored fibers. The physical separation between *GhTT2-3A* and *GbBM* spans 9.2 MB, whereas the gap between *GhTT2-3A* and *PCC/GhTT19* measures 7.6 MB (Figure 1B and Supplementary Table 2). The genes *GbBM*, *GhTT2-3A*, and

PCC/GhTT19 are closely located on chromosome A07 in cotton, suggesting that the genes responsible for the coloration of cotton reproductive organs likely constitute a metabolic gene cluster at the distal end of chromosome A07.

Further examining the color-determining components, we found that the anthocyanin levels in *pcc* cotton petals, including pelargonidin, cyanidin, and delphinidin, were significantly higher than in white-petaled cotton Y18R, both during and post-flowering. Notably, the pronounced accumulation of cyanidin in *pcc* cotton emerged as the key factor for its red flower phenotype (Figures 1C, D). Similarly, the presence of cyanidin, delphinidin, and petunidin was instrumental in forming basal spots on the petals of *GbBM*-induced upland cotton. Additionally, the brown hue of the fibers in brown cotton was predominantly due to the synthesis and buildup of proanthocyanidins.

Genetic variation and population structure of the distal region of chromosome A07 in cotton

To analyze the genetic variation of the distal region of chromosome A07 in cotton, we utilized resequencing data from 98 cotton accessions, including 60 *G. hirsutum* (AD1), 19 *G. arboreum* (A2), and 19 *G. barbadense* (AD2). These data were mapped to the reference genome of *G. hirsutum* (TM-1). We identified a total of 120,126 high-quality SNPs. Population structure analysis and construction of a phylogenetic tree were conducted to understand the genetic diversity within populations. The 98 accessions were classified into four subgroups (G1, G2, G3, and G4) when the value of k was set to 4 (Figure 2A and Supplementary Table 4). G1 subgroup primarily consisted of *G. hirsutum* with white fiber, G2 subgroup predominantly included *G. hirsutum* with brown fiber, G3 subgroup mainly represented individuals of *G. barbadense*, and G4 subgroup was mainly composed of *G. arboreum*. The results from the phylogenetic tree were consistent with results of population genetic structure analysis (Figures 2A, B and Supplementary Table 4).

The nucleotide diversity (π) of G1 (0.00038) and G2 (0.00049) subgroups was lower compared to G3 (0.0044) and G4 (0.00087) (Figure 2C). Both G1 and G2 subgroup belonged to *G. hirsutum*, while G3 and G4 subgroup were different species. By comparing the *Fst* values between each group, we observed that the *Fst* between G1 and G2 subgroup was the lowest, while the differences in *Fst* between the other two groups were more significant. These results suggest that upland cotton generally has lower genetic diversity and a more stable population. Furthermore, the data indicated that G2 subgroup had higher genetic diversity than G1 subgroup, indicating that brown-fibered upland cotton maintains greater genetic diversity compared to white-fibered upland cotton. Additionally, the results revealed that prolonged artificial selection for long-fibered upland cotton varieties has gradually reduced fiber color diversity, resulting in decreased genetic diversity in modern cultivated white-fibered upland cotton.
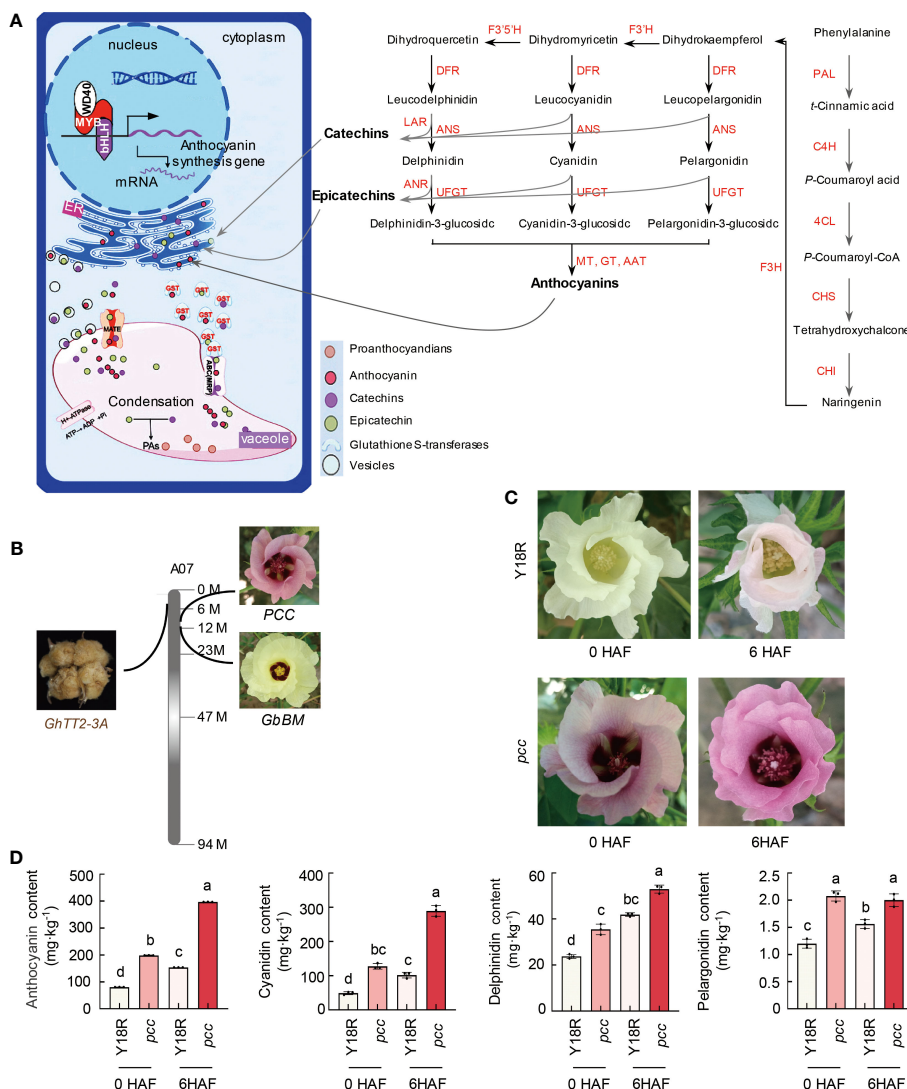
FIGURE 1

The identification of color genes and a model for anthocyanin biosynthesis and transport. **(A)** The model illustrates the synthesis and transport pathway of anthocyanins and procyanidins in plants. PAL, phenylalanine ammonia-lyase. C4H, cinnamate-4-hydroxylase. 4CL, 4-coumarate-CoA ligase. CHS, chalcone synthase. CHI, chalcone-flavonone isomerase. F3'5'H, flavonoid 3'-monooxygenase. F3'H, flavanone 3ß-hydroxylase. ANR, anthocyanidin reductase. LAR, leucoanthocyanidin reductase. DFR, dihydroflavonol-4-reductase. ANS, anthocyanidin synthase. UFGT, UDP-glucose flavonoid glycosyltransferase. **(B)** Localization of genes related to the color of cotton's reproductive organs on the distal region of chromosome A07. PCC/GhTT19 is identified as the gene controlling the red color of G. hirsutum flowers. GbBM is highlighted as a crucial gene influencing the presence of purple spots on G. barbadense flowers. GhTT2-3A is recognized as the primary gene responsible for the brown color of cotton fibers. **(C)** The flower color of Y18R and pcc during and after the flowering stage. HAF, hours after flowering. **(D)** Anthocyanin content during and after the flowering period. Data are shown as mean ± SD from three replicates. Statistically significant differences (P<0.05) are indicated by different letters, based on one-way ANOVA analysis. HAF, hours after flowering.

Although the F_{st} difference between G1 and G2 subgroup is relatively low (Figure 2C), we still observed some distinct variations in the genome-wide comparison. Notably, there is a significant peak on chromosome A07 (Figure 2D). Subsequently, we conducted an analysis specifically focusing on the chromosome A07 and found a distinct difference in the distal region of chromosome A07 (Figure 2E). This observation suggests that distal region of chromosome A07 may play a crucial role in determining the brown fibers in cotton. These findings align with previous research that has identified genes responsible for controlling cotton fiber color on the chromosome A07 (Wen et al., 2018).

Analysis of the GST gene family copy number and expression in the distal region of cotton chromosome A07: implications for anthocyanin transport

Gene mapping was conducted on GST genes in 8 different cotton varieties, namely, *G. herbaceum* (A1), *G. arboretum* (A2), *G. raimondii* (D5), *G. hirsutum* (AD1), *G. barbadense* (AD2), *G. mustelinum* (AD3), *G. tomentosum* (AD4), *G. darwinii* (AD5), revealing their distribution across all chromosomes in different cotton species. We identified 70, 59, 73, 110, 113, 134, 157, and 149
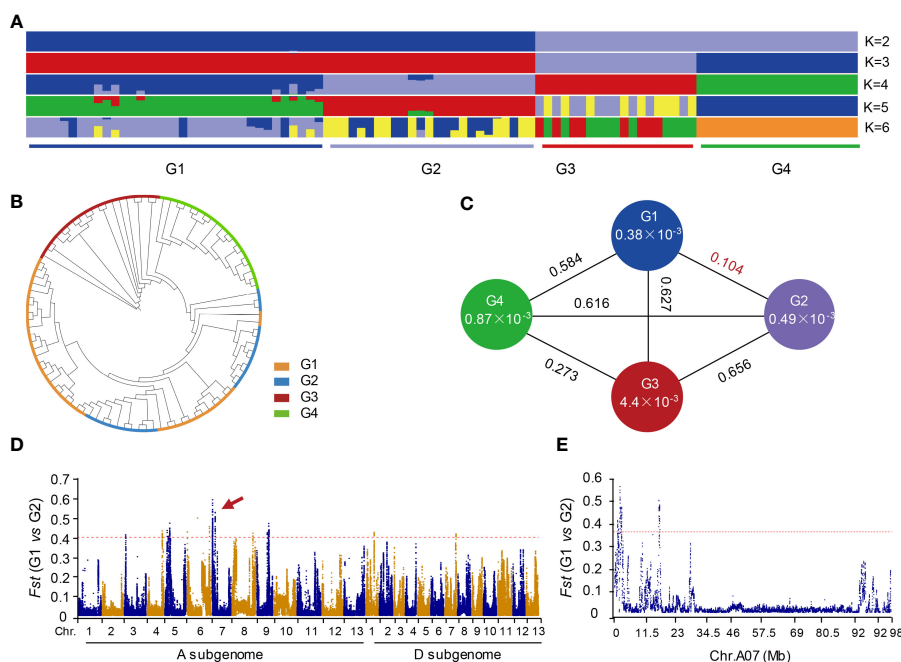
FIGURE 2

Phylogenetic relationships and genetic differentiation among cotton accessions through a comprehensive analysis. **(A)** The population structure of cotton accessions is depicted, with the accessions categorized into four distinct groups at K = 4. Group G1 comprises *G. hirsutum* accessions characterized by white fiber, while G2 includes *G. hirsutum* accessions with brown fiber. Group G3 is represented by *G. arboreum*, and G4 consists of *G. barbadense* accessions. **(B)** A phylogenetic tree constructed using 120,126 high-quality single nucleotide polymorphisms (SNPs). **(C)** Genetic diversity and population differentiation among the four groups. The nucleotide diversity (π) value for each group is indicated within the respective circles. The values between each pair of groups denote the population divergence index (*Fst*). **(D)** Pairwise comparisons of *Fst* values between groups G1 and G2. G1 is *G. hirsutum* with the white fiber. G2 is *G. hirstum* with the brown fiber. **(E)** Pairwise comparisons of *Fst* values between G1 and G2 on chromosome A07.

GST genes in these cotton varieties, respectively. Interestingly, the number of GST gene family members in tetraploid cotton is approximately twice that in diploid cotton (Supplementary Figure 2A), it reveals that genome duplication results in a twofold increase in the number of genes. Notably, chromosome Chr03 exhibited the highest number of tandemly duplicated gene clusters for GST genes, while chromosome 06 had the fewest in all diploid cotton species (Supplementary Figures 3, 4). However, in tetraploid cotton, a significant number of GST genes were found on chromosome A02 and D02, with some forming tandemly duplicated gene clusters (Supplementary Figures 5-9). In contrast, chromosome A07 contained only four GST genes in both diploid and tetraploid cotton species, and no gene clusters were observed resulting from tandem duplication of GST genes (Figure 3A). The number and location of GST genes on chromosome A07 are consistent throughout the evolution of cotton from diploid to tetraploid, indicating that the gene count remains unchanged during the whole genome duplication (WGD) and tandem duplication events. This result emphasizes the conservative nature of GST genes on the chromosome A07.

Subsequent synteny analysis of the GST gene family revealed a high collinearity of GST genes on chromosome A07 in *G. herbaceum*, *G. hirsutum*, and *G. barbadense*, suggesting their likely origin from *G. herbaceum* (Supplementary Figure 19). In *G. hirsutum* and *G. barbadense*, homologous GST genes on chromosomes A02 and A03 exhibited positional changes compared to *G. herbaceum*,

indicating a replacement of chromosomes A02 and A03 in the A subgenome during the diploid to tetraploid transition (Figure 3B). Furthermore, synteny analysis between tetraploid cotton species and *G. arboreum* with *G. herbaceum* revealed a higher homology of the A subgenome of all tetraploid cotton species with *G. herbaceum* (Supplementary Figure 19).

Observations from the constructed phylogenetic tree reveal that *Ghe07G08770*, *Garb_07G021300*, *Ghir_A07G008080*, *Gbar_A07G007920*, *Godar.A07G086400*, *Gotom.A07G086000*, *Gomus.A07G083400* cluster together in the Phi subfamily (Supplementary Figure 21 and Supplementary Table 5). Previous studies have indicated that plant GST genes in the Phi subfamily are involved in the transport and accumulation of flavonoids (Kitamura et al., 2012). Sequence alignments demonstrate a high degree of conservation in their sequences, with only two mutation sites identified in *G. herbaceum* (Supplementary Figure 22), further supporting the notion of high conservation of GST genes related to anthocyanin accumulation and transport across different cotton species. Real-time PCR and RNA-seq analysis to investigate the expression patterns of GST genes in the *pcc* and Y18R of *G. hirsutum* during flowering time, reveal that the expression of only the *Ghir_A07G008080* (*PCC*) gene on the chromosome A07 is detected in flowers. Its expression in *pcc* is significantly higher than in the control cotton Y18R, reaching 90 times the expression level of Y18R (Figures 3C, D and Supplementary Figure 17) The expression of the other three genes is barely detectable (Supplementary
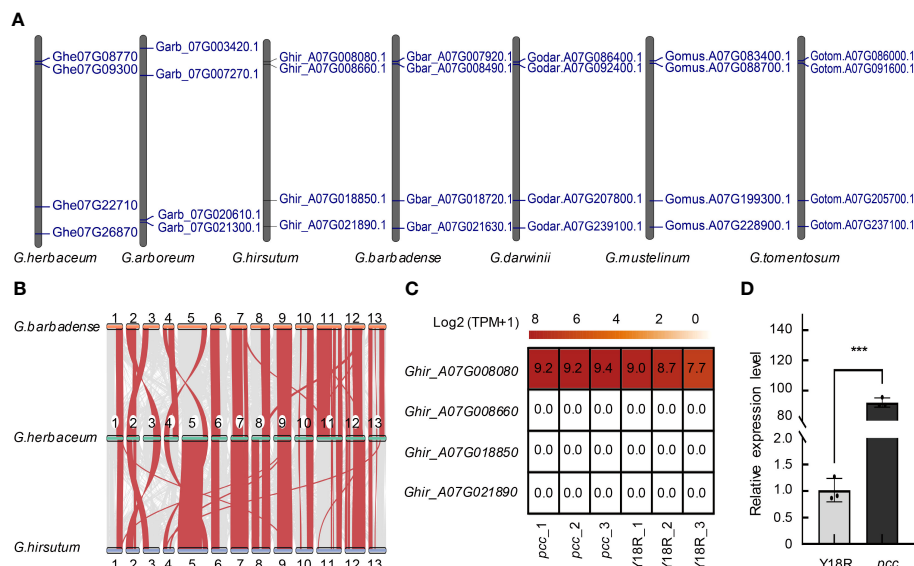
FIGURE 3

The distribution, Synteny, and expression of the GST gene. **(A)** The distribution of GST gene family members across chromosome A07 or 07 in seven cotton varieties. **(B)** The synteny analysis of GST genes among *G. herbaceum*, *G. hirsutum*, and *G. barbadense*. **(C)** The expression patterns of GST genes located on chromosome A07 during the flowering stages. The analysis is based on data from three biological replicates for each group. **(D)** Validation of *Ghir_A07G008080* expression in Y18R and *pcc*. Data are presented as mean ± SD from three replicates. Statistical significance is denoted by asterisks: * *p*<0.05; ** *p*<0.01; *** *p*<0.001, determined by t-test.

Figure 17C). It further confirms the key gene of *Ghir_A07G008080* influencing the formation of different flower colors in *G. hirsutum*.

Exploring the MYB TF family in the distal region of cotton chromosome A07: implications for the anthocyanin pathway

To gain insights into the regulation of reproductive organ color by MYB transcription factors located on the distal end of chromosome A07 in cotton, we conducted a comprehensive analysis. Firstly, we identified the MYB genes present in various cotton species. We respectively identified 216, 211, 218, 416, 426, 430, 438, and 429 MYB genes in the eight cotton varieties (Supplementary Figure 2B). Subsequently, we plotted their genomic locations on a chromosomal map, providing a visual representation of their distribution. The distribution of MYB genes across chromosomes in different cotton species was found to be random and uneven. Specifically, in diploid cotton, the MYB genes were predominantly located on chromosomes 05, 07, 08, 11, 12, and 13. Similarly, in tetraploid cotton, the majority of MYB genes in the A subgenome were found on chromosomes A05, A07, A08, A11, A12, and A13 (Supplementary Figures 10-16).
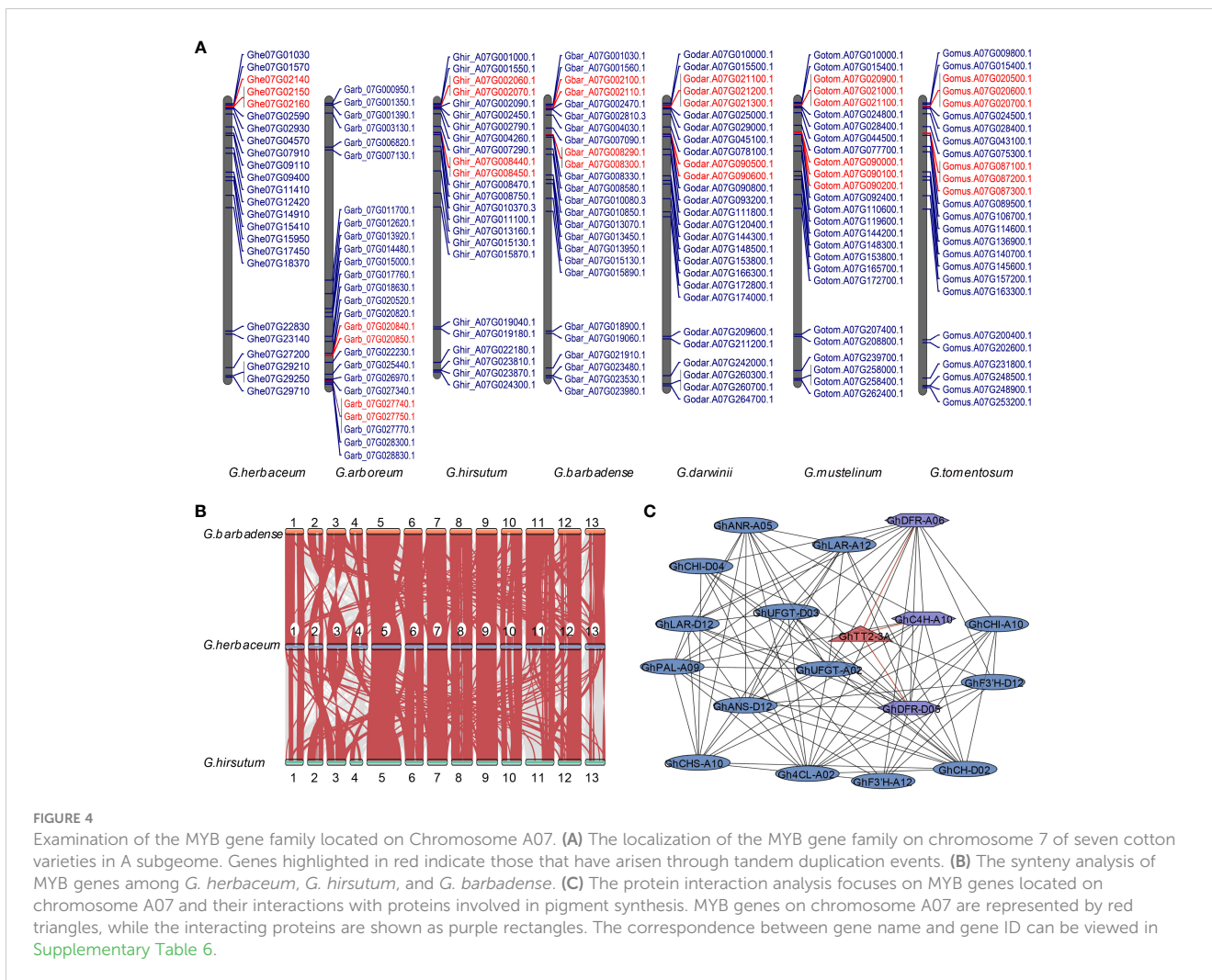
We made some interesting observations in our study. Firstly, we found tandem repeat gene clusters on both chromosome 07 in diploid cotton and A07 in tetraploid cotton (Figure 4A), except in *G. herbaceum*, where only one tandem repeat gene cluster was present on chromosome A07. Secondly, during synteny analysis of the MYB gene family, we discovered a high degree of homology between chromosome 07 in *G. herbaceum* and the A07 chromosome in *G. hirsutum* and *G. barbadense* (Figure 4B). Next,

we observed the expression levels of MYB genes on the A07 chromosome in white and brown cotton at 0, 5, 10, 15, and 20 days post-anthesis (DPA). It was evident that only the *Ghir_A07G002090* gene showed significant differences between white and brown cotton fibers. The differences were most pronounced at 5, 10, and 15 DPA, reaching approximately 2-3 times higher expression levels in white cotton fibers (Supplementary Figure 18). Lastly, the synteny analysis revealed that the A07 chromosome remained evolutionarily conserved during the transition from diploid to tetraploid, showing no significant variation (Supplementary Figure 20). This conservation is likely attributed to the importance of maintaining stability in the anthocyanin metabolism process, thereby preventing substantial changes.

Our protein-protein interaction analysis focused on the interactions between MYB proteins on the chromosome A07 and proteins involved in the anthocyanin and proanthocyanin pathways. For example, we observed that *GhTT2-3A* showed interactions with proteins related to anthocyanin synthesis, such as GhDFR_A06, GhC4H_A10 and GhDFR_D05. (Figure 4C and Supplementary Table 6). This finding further suggests that MYB TF in the distal region of cotton chromosome A07 plays a crucial role in the anthocyanin metabolism process and likely influences the coloration of brown cotton fibers.

Discussion

In this study, we pinpointed two MYB transcription factor genes, *GbBM* and *GhTT2-3A*, which are pivotal in anthocyanin synthesis, influencing the purple spots at petal bases and the brown

FIGURE 4

Examination of the MYB gene family located on Chromosome A07. **(A)** The localization of the MYB gene family on chromosome 7 of seven cotton varieties in A subgeome. Genes highlighted in red indicate those that have arisen through tandem duplication events. **(B)** The synteny analysis of MYB genes among *G. herbaceum*, *G. hirsutum*, and *G. barbadense*. **(C)** The protein interaction analysis focuses on MYB genes located on chromosome A07 and their interactions with proteins involved in pigment synthesis. MYB genes on chromosome A07 are represented by red triangles, while the interacting proteins are shown as purple rectangles. The correspondence between gene name and gene ID can be viewed in Supplementary Table 6.

hue of fibers, respectively. Alongside, we identified the *PCC/GhTT19* gene, crucial for anthocyanin transport to vacuoles, thereby dictating petal redness, located in the distal region of chromosome A07. These genes collectively form a cluster that orchestrates the coloration in cotton's reproductive structures. Our population genetics analysis revealed that significant genetic variability in chromosome A07's distal region plays a crucial role in determining fiber brown coloration. Moreover, the pronounced conservation observed at this chromosome's distal end underscores its role in ensuring anthocyanin metabolic stability, potentially boosting offspring's reproductive success.

The finding that non-homologous genes with similar functions can form metabolic gene clusters highlights a key evolutionary process, distinct from bacteria's horizontal gene transfer (Richard et al., 2008; Sue et al., 2011; Takos et al., 2011). In eukaryotes, these clusters likely result from gene duplication, neofunctionalization, and dynamic genomic rearrangements, a departure from bacterial operon structures where genes are co-transcribed. Eukaryotic cluster genes are transcribed separately, indicating a sophisticated regulatory system (Nutzmann et al., 2018). Whole-genome duplication (WGD) is a critical evolutionary milestone, leading to

rapid diploidization as redundant genes are eliminated to preserve genomic stability, though some duplicates remain (Freeling, 2009; Li et al., 2021). These duplications are fundamental to evolutionary development.

Plant gene clusters can arise through tandem repeats, as seen in the maize TPS gene cluster, or through dispersed repeats facilitated by transposons, like in the poppy (Winzer et al., 2012; Ding et al., 2020). These clusters typically locate in chromosomal hotspots such as telomeres and centromeres, areas prone to recombination, rearrangement, and translocation, promoting gene cluster diversification and formation (Field et al., 2011; Field and Osbourn, 2012; Zhou et al., 2022). Yet, the regulatory mechanisms governing plant metabolic gene clusters remain largely elusive, with scant examples at the transcriptional and chromatin levels. For instance, in Oryza sativa, the OsTGAP1 transcription factor, triggered by chitin oligosaccharides, controls a diterpenes gene cluster, influencing diterpene production (Okada et al., 2009). Similarly, in *Arabidopsis thaliana*, a triterpene gene cluster undergoes spatial conformational changes in its structural domains upon activation or suppression, underscoring a sophisticated regulatory framework (Nutzmann et al., 2020).

Reproductive organ coloration is vital for attracting pollinators and enhancing plant resilience, driven by the evolutionary interaction between plants and pollinators (Zhu et al., 2017; Su et al., 2020). This interaction has led to the development of flower colors and patterns that maximize visibility and attractiveness according to pollinator preferences (Chittka and Raine, 2006). For example, *G. hirsutum*'s yellow petals with purple spots not only provide UV protection but also enhance visibility for pollinators, and its yellow petals are particularly attractive to bees (Abid et al., 2023). In contrast, modern upland cotton has developed milky-white, unspotted flowers, likely due to selective breeding aimed at minimizing dependence on insect pollinators, leading to a gradual loss of distinctive coloration and patterns (Zhang et al., 2016; Cai et al., 2023). This coloration is closely associated with anthocyanin metabolism, essential for the plant's anti-aging, antioxidant, and stress resistance capabilities (He and Giusti, 2010; Alappat and Alappat, 2020). The coloring process is regulated by transcription factors that control anthocyanin biosynthesis genes and transport proteins that facilitate pigment accumulation (Marrs et al., 1995; Quattrocchio et al., 1999; Sun et al., 2012; Ni et al., 2023). Our research indicates that the red petals of *G. hirsutum pcc* and the purple spots on *G. barbadense* petals result from anthocyanin accumulation, triggered by specific gene activity. Additionally, the brown fiber coloration in cotton is due to the upregulated expression of the *GhTT2-3A* gene, which boosts proanthocyanidin biosynthesis and accumulation, key to the plant's pigment metabolism.

We have shown that *GhTT2-3A*, *GbBM*, and *PCC/GhTT19* form a cluster at the distal end of chromosome A07. This arrangement implies a synergistic effect on the metabolic processes of anthocyanin and proanthocyanidin. The location of these genes near the telomeric region suggests that gene rearrangements and exchanges during population differentiation might have played a role in bringing together genes that regulate both anthocyanin and flavonoid biosynthesis and transport within this specific genomic area. This clustering is believed to boost gene functionality, facilitate regulatory interactions among the genes, and enhance the efficiency of enzymatic reactions within the cluster (Rocha, 2008; Shen et al., 2021). Moreover, this part of the genome shows high stability in tetraploid cotton, indicating that the initial grouping of these genes was vital for preserving the stability of anthocyanin and pigment metabolism. These findings provide valuable insights into the genetic mechanisms behind color variation in cotton reproductive organs, deepening our understanding of the complex regulatory networks that control pigment biosynthesis.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

LZ: Writing – original draft, Writing – review & editing, Software, Methodology. JZ: Methodology, Software, Writing – review & editing. HH: Methodology, Writing – review & editing. ZM: Methodology, Writing – review & editing. YW: Data curation, Investigation, Resources, Writing – review & editing. SG: Data curation, Methodology, Writing – review & editing. CL: Supervision, Writing – original draft, Writing – review & editing, Conceptualization, Project administration.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1381071/full#supplementary-material

References

Abid, M. A., Wei, Y., Meng, Z., Wang, Y., Ye, Y., Wang, Y., et al. (2022). Increasing floral visitation and hybrid seed production mediated by beauty mark in *Gossypium hirsutum*. *Plant Biotechnol. J.* 20, 1274–1284. doi: 10.1111/pbi.13805

Abid, M. A., Zhou, Q., Abbas, M., He, H., Meng, Z., Wang, Y., et al. (2023). Natural variation in Beauty Mark is associated with UV-based geographical adaptation in *Gossypium* species. *BMC Biol.* 21, 106. doi: 10.1186/s12915-023-01591-5

Alappat, B., and Alappat, J. (2020). Anthocyanin pigments: Beyond aesthetics. *Molecules* 25, 5500. doi: 10.3390/molecules25235500

Belwal, T., Singh, G., Jeandet, P., Pandey, A., Giri, L., Ramola, S., et al. (2020). Anthocyanins, multi-functional natural products of industrial relevance: Recent biotechnological advances. *Biotechnol. Adv.* 43, 107600. doi: 10.1016/j.biotechadv.2020.107600

Buchfink, B., Reuter, K., and Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. doi: 10.1038/s41592-021-01101-x

Cai, C., Zhou, F., Li, W., Yu, Y., Guan, Z., Zhang, B., et al. (2023). The R2R3-MYB transcription factor *GaPC* controls petal coloration in cotton. *Crop J.* 11, 1319–1330. doi: 10.1016/j.cj.2023.03.013

Cardenas, P. D., Sonawane, P. D., Pollier, J., Vanden Bossche, R., Dewangan, V., Weithorn, E., et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* 7, 10654. doi: 10.1038/ncomms10654

Chai, Q., Wang, X., Gao, M., Zhao, X., Chen, Y., Zhang, C., et al. (2023). A glutathione S-transferase *GhTT19* determines flower petal pigmentation via regulating anthocyanin accumulation in cotton. *Plant Biotechnol. J.* 21, 433–448. doi: 10.1111/pbi.13965

Chittka, L., and Raine, N. E. (2006). Recognition of flowers by pollinators. *Curr. Opin. Plant Biol.* 9, 428–435. doi: 10.1016/j.pbi.2006.05.002

Chu, H. Y., Wegel, E., and Osbourn, A. (2011). From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J.* 66, 66–79. doi: 10.1111/j.1365-313X.2011.04503.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008

Ding, Y., Weckwerth, P. R., Poretsky, E., Murphy, K. M., Sims, J., Saldivar, E., et al. (2020). Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nat. Plants* 6, 1375–1388. doi: 10.1038/s41477-020-00787-9

Duan, Q., Goodale, E., and Quan, R. (2014). Bird fruit preferences match the frequency of fruit colours in tropical Asia. *Sci. Rep.* 4, 5627. doi: 10.1038/srep05627

Field, B., Fiston-Lavier, A. S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16116–16121. doi: 10.1073/pnas.1109273108

Field, B., and Osbourn, A. (2012). Order in the playground: Formation of plant gene clusters in dynamic chromosomal regions. *Mol. Genet. Elements* 2, 46–50. doi: 10.4161/mge.19348

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Gomez, C., Conejero, G., Torregrosa, L., Cheynier, V., Terrier, N., and Ageorges, A. (2011). *In vivo* grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J.* 67, 960–970. doi: 10.1111/j.1365-313X.2011.04648.x

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, 1178–1186. doi: 10.1093/nar/gkr944

Grotewold, E., and Davies, K. (2008). Trafficking and sequestration of anthocyanin. *Nat. Prod Commun.* 3, 1251–1258. doi: 10.1177/1934578X0800300806

Guo, L., Qiu, J., Ye, C., Jin, G., Mao, L., Zhang, H., et al. (2017). Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat. Commun.* 8, 1031. doi: 10.1038/s41467-017-01067-5

He, J., and Giusti, M. M. (2010). Anthocyanins: natural colorants with health-promoting properties. *Annu. Rev. Food Sci. Technol.* 1, 163–187. doi: 10.1146/annurev.food.080708.100754

He, G., Zhang, R., Jiang, S., Wang, H., and Ming, F. (2023). The MYB transcription factor RcMYB1 plays a central role in rose anthocyanin biosynthesis. *Hortic. Res.* 10, uhad080. doi: 10.1093/hr/uhad080

He, D., Zhao, X., Liang, C., Zhu, T., Abid, M. A., Cai, Y., et al. (2018). Genetic variation in LBL1 contributes to depth of leaf blades lobes between cotton subspecies, Gossypium barbadense and *Gossypium hirsutum*. *J. Integr. Agr* 17, 2394–2404. doi: 10.1016/S2095-3119(18)61954-9

Hu, K. (2021). Become Competent in Generating RNA-seq heat maps in one day for novices without prior r experience. *Methods Mol. Biol.* 2239, 269–303. doi: 10.1007/978-1-0716-1084-8_17

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A. J., Shinde, B., Cardenas, P. D., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 341, 175–179. doi: 10.1126/science.1240230

Jonczyk, R., Schmidt, H., Osterrieder, A., Fiesselmann, A., Schullehner, K., Haslbeck, M., et al. (2008). Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. *Plant Physiol.* 146, 1053–1063. doi: 10.1104/pp.107.111237

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kitamura, S., Akita, Y., Ishizaka, H., Narumi, I., and Tanaka, A. (2012). Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. *J. Plant Physiol.* 169, 636–642. doi: 10.1016/j.jplph.2011.12.011

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, Z., McKibben, M. T. W., Finch, G. S., Blischak, P. D., Sutherland, B. L., and Barker, M. S. (2021). Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.* 72, 387–410. doi: 10.1146/annurev-arplant-050718-100344

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, Z., Cheema, J., Vigouroux, M., Hill, L., Reed, J., Paajanen, P., et al. (2020). Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat. Commun.* 11, 5354. doi: 10.1038/s41467-020-19153-6

Liu, W., Feng, Y., Yu, S., Fan, Z., Li, X., Li, J., et al. (2021). The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* 22, 12824. doi: 10.3390/ijms222312824

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Loyall, L., Uchida, K., Braun, S., Furuya, M., and Frohnmeyer, H. (2000). Glutathione and a UV light-induced glutathione S-transferase are involved in signaling to chalcone synthase in cell cultures. *Plant Cell* 12, 1939–1950. doi: 10.1105/tpc.12.10.1939

Luan, Y., Tang, Y., Wang, X., Xu, C., Tao, J., and Zhao, D. (2022). Tree Peony R2R3-MYB Transcription factor PsMYB30 promotes petal blotch formation by activating the transcription of the anthocyanin synthase gene. *Plant Cell Physiol.* 63, 1101–1116. doi: 10.1093/pcp/pcac085

Ma, H., Yang, T., Li, Y., Zhang, J., Wu, T., Song, T., et al. (2021). The long noncoding RNA *MdLNC499* bridges *MdWRKY1* and *MdERF109* function to regulate early-stage light-induced anthocyanin accumulation in apple fruit. *Plant Cell* 33, 3309–3330. doi: 10.1093/plcell/koab188

Marrs, K. A., Alfenito, M. R., Lloyd, A. M., and Walbot, V. (1995). A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. *Nature* 375, 397–400. doi: 10.1038/375397a0

Ni, J., Wang, S., Yu, W., Liao, Y., Pan, C., Zhang, M., et al. (2023). The ethylene-responsive transcription factor PpERF9 represses *PpRAP2.4* and *PpMYB114* via histone deacetylation to inhibit anthocyanin biosynthesis in pear. *Plant Cell* 35, 2271–2292. doi: 10.1093/plcell/koad077

Nutzmann, H. W., Doerr, D., Ramirez-Colmenero, A., Sotelo-Fonseca, J. E., Wegel, E., Di Stefano, M., et al. (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. U.S.A.* 117, 13800–13809. doi: 10.1073/pnas.1920474117

Nutzmann, H. W., Scazzocchio, C., and Osbourn, A. (2018). Metabolic gene clusters in eukaryotes. *Annu. Rev. Genet.* 52, 159–183. doi: 10.1146/annurev-genet-120417-031237

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., et al. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J. Biol. Chem.* 284, 26510–26518. doi: 10.1074/jbc.M109.036871

Otasek, D., Morris, J. H., Boucas, J., Pico, A. R., and Demchak, B. (2019). Cytoscape automation: empowering workflow-based network analysis. *Genom Biol.* 20, 185. doi: 10.1186/s13059-019-1758-4

Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 181, 219–229. doi: 10.1016/j.plantsci.2011.05.009

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Pucker, B., and Selmar, D. (2022). Biochemistry and molecular basis of intracellular flavonoid transport in plants. *Plants* 11, 963. doi: 10.3390/plants11070963

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., et al. (1999). Molecular analysis of the *anthocyanin 2* gene of petunia and its role in the evolution of flower color. *Plant Cell* 11, 1433–1444. doi: 10.1105/tpc.11.8.1433

Retief, J. D. (2000). Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* 132, 243–258. doi: 10.1385/1-59259-192-2:243

Richard, G. F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72, 686–727. doi: 10.1128/MMBR.00011-08

Rocha, E. P. (2008). The organization of the bacterial genome. *Annu. Rev. Genet.* 42, 211–233. doi: 10.1146/annurev.genet.42.110807.091653

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Shen, S., Peng, M., Fang, H., Wang, Z., Zhou, S., Jing, X., et al. (2021). An Oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci. Bull.* 66, 2369–2380. doi: 10.1016/j.scib.2021.03.015

Shi, L., Chen, X., Wang, K., Yang, M., Chen, W., Yang, Z., et al. (2021). MrMYB6 from Chinese bayberry (*Myrica rubra*) negatively regulates anthocyanin and proanthocyanidin accumulation. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.685654

Shi, L., Li, X., Fu, Y., and Li, C. (2023). Environmental stimuli and phytohormones in anthocyanin biosynthesis: A comprehensive review. *Inter J. Mol. Sci.* 24, 16415. doi: 10.3390/ijms242216415

Stintzing, F. C., and Carle, R. (2004). Functional properties of anthocyanins and betalains in plants, food, and in human nutrition. *Trends Food Sci. Technol.* 15, 19–38. doi: 10.1016/j.tifs.2003.07.004

Su, W., Tao, R., Liu, W., Yu, C., Yue, Z., He, S., et al. (2020). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* 18, 479–490. doi: 10.1111/pbi.13213

Sue, M., Nakamura, C., and Nomura, T. (2011). Dispersed benzoxazinone gene cluster: molecular characterization and chromosomal localization of glucosyltransferase and glucosidase genes in wheat and rye. *Plant Physiol.* 157, 985–997. doi: 10.1104/pp.111.182378

Sun, Y., Li, H., and Huang, J. R. (2012). *Arabidopsis TT19* functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol. Plant* 5, 387–400. doi: 10.1093/mp/ssr110

Sunil, L., and Shetty, N. P. (2022). Biosynthesis and regulation of anthocyanin pathway genes. *Appl. Microbiol. Biotechnol.* 106, 1783–1798. doi: 10.1007/s00253-022-11835-z

Takos, A. M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., Motawia, M. S., et al. (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *Plant J.* 68, 273–286. doi: 10.1111/j.1365-313X.2011.04685.x

Tan, J., Wang, M., Tu, L., Nie, Y., Lin, Y., and Zhang, X. (2013). The flavonoid pathway regulates the petal colors of cotton flower. *PloS One* 8, e72364. doi: 10.1371/journal.pone.0072364

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the genome analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43, 11 10 11–11 10 33. doi: 10.1002/0471250953.bi1110s43

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wen, T., Wu, M., Shen, C., Gao, B., Zhu, D., Zhang, X., et al. (2018). Linkage and association mapping reveals the genetic basis of brown fibre (*Gossypium hirsutum*). *Plant Biotechnol. J.* 16, 1654–1666. doi: 10.1111/pbi.12902

Winkel, B. S. J. (2004). Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 55, 85–107. doi: 10.1146/annurev.arplant.55.031903.141714

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485

Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., et al. (2012). A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 336, 1704–1708. doi: 10.1126/science.1220757

Xu, Y., Simon, J. E., Ferruzzi, M. G., Ho, L., Pasinetti, G. M., and Wu, Q. (2012). Quantification of anthocyanidins in the grapes and grape juice products with acid assisted hydrolysis using LC/MS. *J. Funct. Foods* 4, 710–717. doi: 10.1016/j.jff.2012.04.010

Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M., et al. (2021). MYB-mediated regulation of anthocyanin biosynthesis. *Inter J. Mol. Sci.* 22, 3103. doi: 10.3390/ijms22063103

Yan, Q., Wang, Y., Li, Q., Zhang, Z., Ding, H., Zhang, Y., et al. (2018). Up-regulation of *GhTT2-3A* in cotton fibres during secondary wall thickening results in brown fibres with improved quality. *Plant Biotechnol. J.* 16, 1735–1747. doi: 10.1111/pbi.12910

Yang, H., Li, H., and Li, Q. (2022). Biosynthetic regulatory network of flavonoid metabolites in stems and leaves of Salvia miltiorrhiza. *Sci. Rep.* 12, 18212. doi: 10.1038/s41598-022-21517-5

Yang, J., Li, B., Shi, W., Gong, Z., Chen, L., and Hou, Z. (2018). Transcriptional activation of anthocyanin biosynthesis in developing fruit of blueberries (*Vaccinium corymbosum* L.) by preharvest and postharvest UV irradiation. *J. Agr Food Chem.* 66, 10931–10942. doi: 10.1021/acs.jafc.8b03081

Yang, Z., Wang, J., Huang, Y., Wang, S., Wei, L., Liu, D., et al. (2023). CottonMD: a multi-omics database for cotton biological study. *Nucleic Acids Res.* 51, D1446–D1456. doi: 10.1093/nar/gkac863

Yu, K., Song, Y., Lin, J., and Dixon, R. A. (2023). The complexities of proanthocyanidin biosynthesis and its regulation in plants. *Plant Commun.* 4, 100498. doi: 10.1016/j.xplc.2022.100498

Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X., et al. (2020). Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat. Plants* 6, 1447–1454. doi: 10.1038/s41477-020-00816-7

Zhang, S., Lan, Q., Gao, X., Yang, B., Cai, C., Zhang, T., et al. (2016). Mapping of genes for flower-related traits and QTLs for flowering time in an interspecific population of *Gossypium hirsutum* × *G. darwinii*. *J. Genet.* 95, 197–201. doi: 10.1007/s12041-016-0617-2

Zhao, J. (2015). Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* 20, 576–585. doi: 10.1016/j.tplants.2015.06.007

Zhao, Y., Jiang, C., Lu, J., Sun, Y., and Cui, Y. (2023). Research progress of proanthocyanidins and anthocyanidins. *Phytother. Res.* 37, 2552–2577. doi: 10.1002/ptr.7850

Zhao, J., Pang, Y., and Dixon, R. A. (2010). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153, 437–443. doi: 10.1104/pp.110.155432

Zhao, S., Ye, Z., and Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* 26, 903–909. doi: 10.1261/rna.074922.120

Zhou, J., Liu, Y., Guo, X., Birchler, J. A., Han, F., and Su, H. (2022). Centromeres: From chromosome biology to biotechnology applications and synthetic genomes in plants. *Plant Biotechnol. J.* 20, 2051–2063. doi: 10.1111/pbi.13875

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., Yang, Z., et al. (2017). Development of "purple endosperm rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Mol. Plant* 10, 918–929. doi: 10.1016/j.molp.2017.05.008