

RESEARCH

Open Access



Genetic factors explaining anthocyanin pigmentation differences

Maria F. Marin-Recinos¹ and Boas Pucker^{1*}

Abstract

Background Anthocyanins are important contributors to coloration across a wide phylogenetic range of plants. Biological functions of anthocyanins span from reproduction to protection against biotic and abiotic stressors. Owing to a clearly visible phenotype of mutants, the anthocyanin biosynthesis and its sophisticated regulation have been studied in numerous plant species. Genes encoding the anthocyanin biosynthesis enzymes are regulated by a transcription factor complex comprising MYB, bHLH and WD40 proteins.

Results A systematic comparison of anthocyanin-pigmented vs. non-pigmented varieties was performed within numerous plant species covering the taxonomic diversity of flowering plants. The literature was screened for cases in which genetic factors causing anthocyanin loss were reported. Additionally, transcriptomic data sets from four previous studies were reanalyzed to determine the genes possibly responsible for color variation based on their expression pattern. The contribution of different structural and regulatory genes to the intraspecific pigmentation differences was quantified. Differences concerning transcription factors are by far the most frequent explanation for pigmentation differences observed between two varieties of the same species. Among the transcription factors in the analyzed cases, *MYB* genes are significantly more prone to account for pigmentation differences compared to *bHLH* or *WD40* genes. Among the structural genes, *DFR* genes are most often associated with anthocyanin loss.

Conclusions These findings support previous assumptions about the susceptibility of transcriptional regulation to evolutionary changes and its importance for the evolution of novel coloration phenotypes. Our findings underline the particular significance of MYBs and their apparent prevalent role in the specificity of the MBW complex.

Keywords Flavonoid biosynthesis, Anthocyanins, Gene expression, Transcription factor, MYB, DFR, Systematic comparison, Transcriptome

*Correspondence:

Boas Pucker

b.pucker@tu-braunschweig.de

¹Plant Biotechnology and Bioinformatics, Institute of Plant Biology and BRICS, TU Braunschweig, Braunschweig, Germany



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Angiosperms are characterized by an enormous diversity of flower hues and shapes [1, 2]. Some plant species maintain their brilliant colors throughout the year, while others constantly transform as the seasons change. The substances responsible for these colors are pigments which include flavonoids, betalains, and carotenoids [3]. These pigment classes differ in their biochemical properties resulting in distinct color ranges. Flavonoids can be classified into multiple subgroups with anthocyanins forming the most colorful subgroup. Anthocyanins can provide orange, red, purple, blue, or almost black coloration [4]. Carotenoids lead to yellow, orange, or red coloration [5]. Betalains can be classified into yellow betaxanthins and red betacyanins [3]. Flavonoid biosynthesis and its regulation are among the best understood processes in plants. This comprehensive understanding makes flavonoids an invaluable system for exploring the mechanistic basis of phenotypic differences in plant coloration. We begin by reviewing the extensive body of knowledge regarding the biochemistry of flavonoid pigmentation before examining trends among the substitutions that commonly contribute to color variation.

Functions of flavonoids

Anthocyanins and other flavonoids, including flavones, flavonols, and proanthocyanidins, are a group of specialized plant metabolites responsible for numerous functions beyond coloration. Additional physiological functions include protection against herbivores [6, 7] and reduction of the impact caused by salinity [8], drought [9], and UV-radiation [10, 11]. Associated with their color are ecological functions such as the attraction of pollinators and seed dispersers which facilitates reproduction [6, 12]. Flavonols also contribute to the attraction of pollinators by forming guiding signals on flowers which are invisible to the human eye [12–15]. Anthocyanins are known to be responsible for the coloration of flowers, with hues ranging from red and orange to purple and blue. The diversity in colors depends on the chemical structure of the anthocyanin compound which includes the number of hydroxyl groups attached to the benzene ring, and the level of glycosylation [16, 17] and acylation [18]. Several reports suggest that the interaction with copigments like flavonols and flavones is an important factor for the stabilization of anthocyanins in plants [19, 20]. These co-pigments can also alter the hue of the anthocyanins participating in the complex formation [21, 22]. Environmental factors can also influence the color stability of anthocyanin pigments. For example, a plant exposed to an acidic soil can produce anthocyanins with an intense red or orange color [23]. Plants exposed to high temperatures can show degradation of

anthocyanins while low temperatures can increase color intensity [24–26].

Flavonols, flavones, and proanthocyanidins have individual biological functions and can influence the coloration of different plant organs [27]. The characteristic cream white or pale yellow color, determined by the presence of flavones and flavonols, can be observed in leaves or petals of *Taraxacum officinale* “dandelion” [28], *Chrysanthemum grandiflorum* cv. Jinba [29], and *Chrysanthemum morifolium* [30]. Proanthocyanidins also called condensed tannins are colorless compounds that turn brown upon oxidation [31]. They have been studied in seeds of species such as *Arabidopsis thaliana* [32], *Brassica napus* [33], and *Ipomoea purpurea* [34].

Structural genes of the flavonoid biosynthesis

The general pathway of the flavonoid biosynthesis (Fig. 1A) is well understood and the central aglycon biosynthesis is largely conserved among land plants [3, 35]. It starts with the condensation of 4-coumaroyl-CoA and malonyl-CoA to synthesize naringenin chalcones which are later isomerized by the enzyme chalcone isomerase (CHI) to form naringenin, a colorless flavanone. In the next step, the pathway diverges: flavanone can either be hydroxylated by the flavanone 3-hydroxylase (F3H) to form dihydroflavonols or it can be oxidized through the activity of flavone synthase (FNS) and form flavones. After the hydroxylation of naringenin to dihydrokaempferol, the formation of dihydroquercetin and dihydromyricetin can take place through the catalysis of flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H), respectively. Subsequently, two enzymes can accept these intermediates and produce either flavonols through oxidation with flavonol synthase (FLS) or leucoanthocyanidins by reduction with dihydroflavonol 4-reductase (DFR). When the latter occurs, another enzyme called anthocyanidin synthase/leucoanthocyanidin dioxygenase (ANS/LDOX) catalyzes the synthesis of anthocyanidins. This step also requires an anthocyanin-related glutathione S-transferase (arGST) that was named AN9/TT19 due to the corresponding mutants [36, 37], but the enzymatic function was only revealed recently [38]. These anthocyanidins can be further modified through different steps including (Fig. 1B) (1) glycosylation in the presence of UDP-glucose flavonoid 3-O-glucosyl transferase (UFGT), (2) methylation through the activity of O-methyltransferase (OMT), and (3) acylation by the anthocyanin acyltransferase (ACT). Moreover, leucoanthocyanidins and anthocyanidins can also be reduced by the enzymatic activity of leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR), respectively, to synthesize catechins and epicatechins leading to the production of proanthocyanidins.

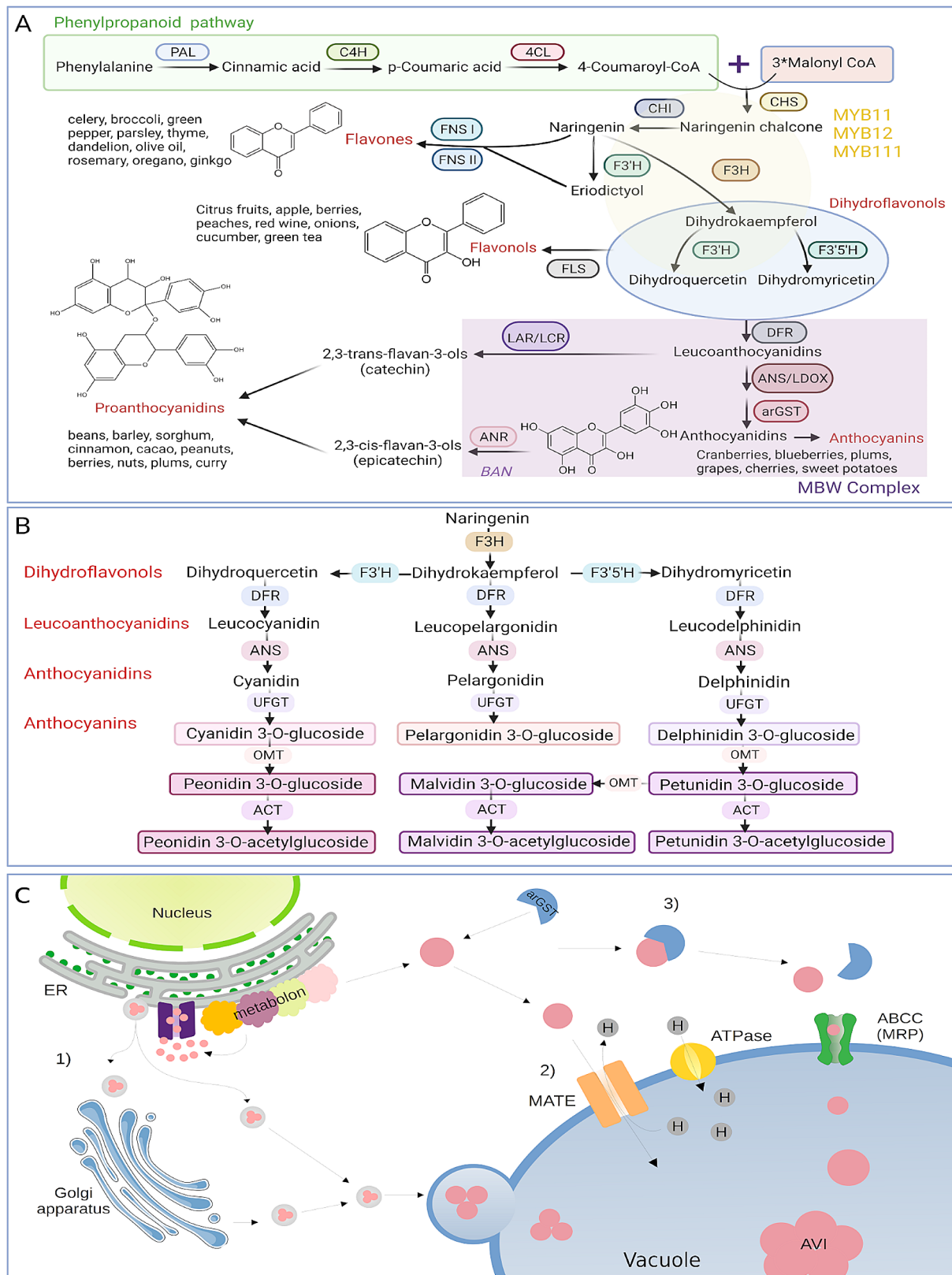


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 (A) Schematic representation of the general flavonoid biosynthesis pathway. (B) Simplified flowchart describing the biosynthesis pathway of anthocyanins. Enzyme names are abbreviated as follows: PAL - phenylalanine ammonia-lyase, C4H - cinnamic acid 4-hydroxylase, 4CL-4-coumarate-CoA ligase, CHS - chalcone synthase, CHI - chalcone isomerase, F3H - flavanone 3-hydroxylase, F3'H - flavonoid 3'-hydroxylase, F3'5'H - flavonoid 3',5'-hydroxylase, DFR - dihydroflavonol 4-reductase, ANS/ LDOX - anthocyanidin synthase /leucoanthocyanidin dioxygenase, arGST - anthocyanin-related glutathione S-transferase, UFGT - UDP-glucose: flavonoid 3-O-glucosyltransferase, FLS - flavonol synthase, ANR - anthocyanidin reductase, FNS - flavone synthase. (C) Scheme of the different intracellular flavonoid transport mechanisms: (1) vesicle trafficking from the Endoplasmic Reticulum (ER), these vesicles may also incorporate Anthocyanin Vacuolar Inclusions (AVI) for efficient transport and storage within the vacuole, (2) membrane transporter throughout the multidrug and toxin extrusion transporter (MATE), and (3) transport might be mediated by the glutathione S-transferase (GST) and the ATP-binding cassette (ABCC) transporter. GST is presented twice due to the recently reported enzymatic function of arGSTs by Eichenberger et al [38]. It is currently not clear if GST functions only as an enzyme or if it also plays a role in anthocyanin transport

Intracellular transportation of flavonoids

Like many other specialized metabolites, flavonoids are synthesized in the cytoplasm. It is assumed that some enzymes responsible for catalyzing specific reactions in the flavonoid biosynthesis are attached to the endoplasmic reticulum (ER) and form a metabolon [39] (Fig. 1C). Flavonoids produced at the ER are transported into the vacuole for storage [40–42] which results in observable pigmentation. The transport of these metabolites to the vacuole is a process that is not fully understood, but different models have been proposed that could explain observations from several experiments. Two widely accepted and well-known models are: (1) vesicle trafficking from the ER to the vacuole and (2) GST-mediated transport to the tonoplast, where membrane-based transporters are active [41]. Both models have in common a mechanism that is required to transport flavonoids across a membrane and these models are not mutually exclusive. It is feasible that these transport routes are active to certain degrees under different conditions, in developmental stages, or in different plant parts. The vesicular transport model proposes the accumulation of flavonoids inside the ER lumen and formation of small flavonoid rich compartments surrounded by a membrane that move to the central vacuole [39, 42]. The existence of these vesicles has been reported mainly in *Zea mays* [43], *A. thaliana* [44], *Vitis vinifera* [45], and *Oryza sativa* [46]. Microscopic evidence shows that these vesicular bodies are attached to the surface of the ER [45, 47] from where they are released into the cytoplasm and mobilized directly into the vacuole either by fusing with carrier proteins, or mobilized indirectly by following the trans-Golgi Network (TGN) transport pathway [48]. The GST-mediated transport model proposes that flavonoids are delivered to the tonoplast by ligandins [36, 49]. These 'ligandins' would be glutathione S-transferase (GST) binding and carrier proteins [50]. Evidence for the role of GST in flavonoid transport has been reported in multiple species, such as *Zea mays* [49], *A. thaliana* [36, 37], *Petunia hybrida* [50], *Vitis vinifera* [45], and *Prunus persica* [51]. Reconsideration of the GST-mediated transporter is needed in the light of a recent study [38] that revealed an enzymatic function in the anthocyanin biosynthesis for anthocyanin-related GSTs (arGSTs). The tonoplast-based

transport mechanism involves different transmembrane channels which enable translocation of flavonoids into the vacuole [39]. These routes were reported to involve the Multidrug Resistance-associated Protein (MRP), belonging to the family of proteins ATP-binding cassette (ABC) actively transporting anthocyanins [52]. Using an electrochemical H^+ gradient to transport substances across membranes, Multidrug and Toxic compound Extrusion (MATE) is considered to regulate the vacuolar sequestration of proanthocyanidin precursors in the seed coat cells [52–54].

Regulation of the flavonoid biosynthesis by transcription factors

The transcriptional activity of genes encoding enzymes of the flavonoid biosynthesis pathway is controlled by numerous transcription factors or even protein complexes comprising of multiple transcription factors (Fig. 1A). Genes of the flavonol and flavone biosynthesis are largely regulated by MYB11, MYB12, and MYB111 [55–57]. The mechanism that regulates the expression of all structural genes in the anthocyanin biosynthesis pathway is commonly known as the MBW complex [58]. The name of the MBW complex is based on the three involved transcription factors: R2R3-MYBs, basic helix-loop-helix (bHLH) proteins, and WD40 proteins. One member of each of the three protein families is required for the complex formation. Different members of the MYB and bHLH family can participate resulting in combinatorial diversity [58]. After the discovery of TT8 in *A. thaliana* [59], Baudry et al. [60]. demonstrated the activity of the MBW complex in regulating the expression of the proanthocyanidins (PA) biosynthesis gene *BANYULS* (*BAN*). The ternary complex responsible for *BAN* regulation in *A. thaliana* is composed of TT2 (MYB123), TT8 (bHLH42), and TTG1 (WD40 family). Years later, it was demonstrated that the anthocyanin biosynthesis is also controlled by MBW complexes [61]. These complexes harbor PAP1, PAP2, MYB113, or MYB114 as the MYB component and GL3 or EGL3 as bHLH component as well as the WD40 protein TTG1 [61].

Activation of the anthocyanin biosynthesis by the MBW complex is evolutionary conserved across angiosperms [62], but the individual components involved can

vary between species [63, 64]. The MBW complex controls other biological processes in plants like trichome formation, root hair development, and proanthocyanidin biosynthesis [65]. Multiple functions lead to different degrees of evolutionary constraints on the individual components of the MBW complex. A number of different MYB partners can participate in the MBW complex and are considered as the specificity determining factor of the MBW complex. For example, PAP1/MYB75, PAP2/MYB90, PAP3/MYB113, PAP4/MYB114 activate the anthocyanin biosynthesis in *Arabidopsis* [66], while TT2/MYB123 would activate the proanthocyanidin biosynthesis [60]. MYB5 was described as another anthocyanin activator in *Fragaria* [64] and proanthocyanidin regulators were reported to activate the anthocyanin biosynthesis in *Vaccinium* [67]. It is also known that different bHLHs can be involved in the MBW complex [68, 69] and that TTG1 can be replaced by LWD1 in strawberry [64]. The DNA binding and protein-protein interaction capacity of MYBs and bHLHs is determined by highly conserved regions [70]. It has been postulated that TTG1 serves as a scaffolding protein that maintains the interaction of MYB and bHLH [71]. This protein-protein interaction involves five WD repeats that account for over 60% of the protein length. Previous reports suggest that the MYB component is most often associated with changes in flower pigmentation indicating low constraints on this component of the MBW complex due to higher functional specialization [72–74].

Evolutionary patterns of anthocyanin pigmentation

Closely related plant species can differ in their anthocyanin repertoire and pigmentation pattern [75, 76]. Such pigmentation differences have frequently been reported between plants of the same species [1, 77–79]. To the best of our knowledge, there are no reports about anthocyanin loss in any major taxonomic lineages except for the replacement of anthocyanins by betalains in some families of the *Caryophyllales* [80–82]. This suggests that anthocyanin pigmentation differences are often intra-specific. All structural genes in the anthocyanin biosynthesis pathway must be functional and active to achieve anthocyanin pigmentation. Mutations in any regulator or enzyme encoding gene of the flavonoid biosynthesis can affect the coloration. Pigmentation differences have been studied in many plant species including *A. thaliana* [83], *Vitis vinifera* [84], *Malus domestica* [85], *Solanum lycopersicum* [86], *Hordeum vulgare* [87], *Nicotiana tabacum* [88], and *Nicotiana glauca* [89]. Substrate competition between different branches of the flavonoid biosynthesis can also have an impact on the anthocyanin accumulation [90]. For example, an increased flavonol production can lead to a pigmentation loss due to insufficient substrate for the anthocyanin biosynthesis [91]. These

natural differences in pigmentation provide an excellent system to study evolutionary processes that lead to the inactivation of a pathway. In theory, a biosynthesis pathway could be interrupted at any of the successive steps [92], but previous research suggests that some genes are more often responsible for pigmentation loss than others [74, 93].

Our main questions revolve around identifying potential “hotspots” in the anthocyanin biosynthesis pathway that frequently lead to pigmentation loss. This loss is not attributed to a higher mutation rate but is rather linked to an increased likelihood of mutation fixation. The relative importance of cis-regulatory and trans-regulatory changes during evolution is of huge interest beyond the flavonoid biosynthesis of plants [94, 95]. MYBs are known to be key regulators in the flavonoid pathway [58] and have been frequently implicated in phenotypic variation in flower color [96–98]. Previous studies suggested that MYB transcription factors are often responsible for evolutionary changes [58, 99, 100]. Our findings will generally help to understand whether biosynthesis pathways in plants are naturally shut off at the first committed step. This would be the logical position when performing metabolic engineering, to avoid substrate being channeled into a dead-end pathway and potentially causing the accumulation of toxic intermediates [101–104]. Given that DFR is often presented as the first committed step of the anthocyanin biosynthesis, it represents a critical point where interruptions can effectively avoid metabolic flux into a dead-end pathway. Specifically, our study addresses three questions: (1) Is the loss of anthocyanin pigmentation within species primarily caused by mutations in transcription factors or structural genes? (2) Is DFR at the start of the anthocyanin biosynthesis branch, more predisposed to be causative for variations in anthocyanin pigmentation loss compared to downstream genes such as *ANS*, *arGST*, or *UFGT*? (3) Do MYBs play a more frequent role in anthocyanin pigmentation loss compared to members of other transcription factor families? To address our research questions, we conducted a systematic intraspecific comparison in numerous flowering plant species, contrasting anthocyanin-pigmented varieties with their non-pigmented counterparts. An extensive literature screening was performed to identify reports of causal genes explaining pigmentation differences between varieties of a plant species or, in rare cases, between closely related (sub)species. In instances where studies lacked conclusive results, we scrutinized data availability and conducted reanalysis when feasible. The quantification of different genes associated with the pigmentation loss supports the crucial role of transcription factors, particularly MYBs.

Results

Taxonomic distribution of analyzed cases across plant families

To evaluate how well the analyzed data sets are distributed across different plant lineages, a taxonomic distribution analysis was performed. The 235 analyzed cases (see Additional file 1: Table S1), cover 53 plant families distributed over 31 orders (Fig. 2). Notably, the order *Rosales*, particularly the *Rosaceae* family, accounted for the highest number of studies (39) showcasing variations

in pigmentation. Second were the *Brassicaceae* family with 24 reported cases, and the *Orchidaceae*, *Fabaceae*, and *Solanaceae* families each contributing 13 cases. Furthermore, the *Ericaceae* family appearing in 9 studies, and the *Asteraceae* and *Liliaceae* families, each appearing in 8 studies, were also noteworthy. While the *Theaceae*, *Poaceae*, and *Paeoniaceae* families were featured in 7 cases each. Other families, including *Lamiaceae*, *Asparagaceae*, *Malvaceae* and *Caryophyllaceae*, were identified in varying frequencies across the reviewed studies.

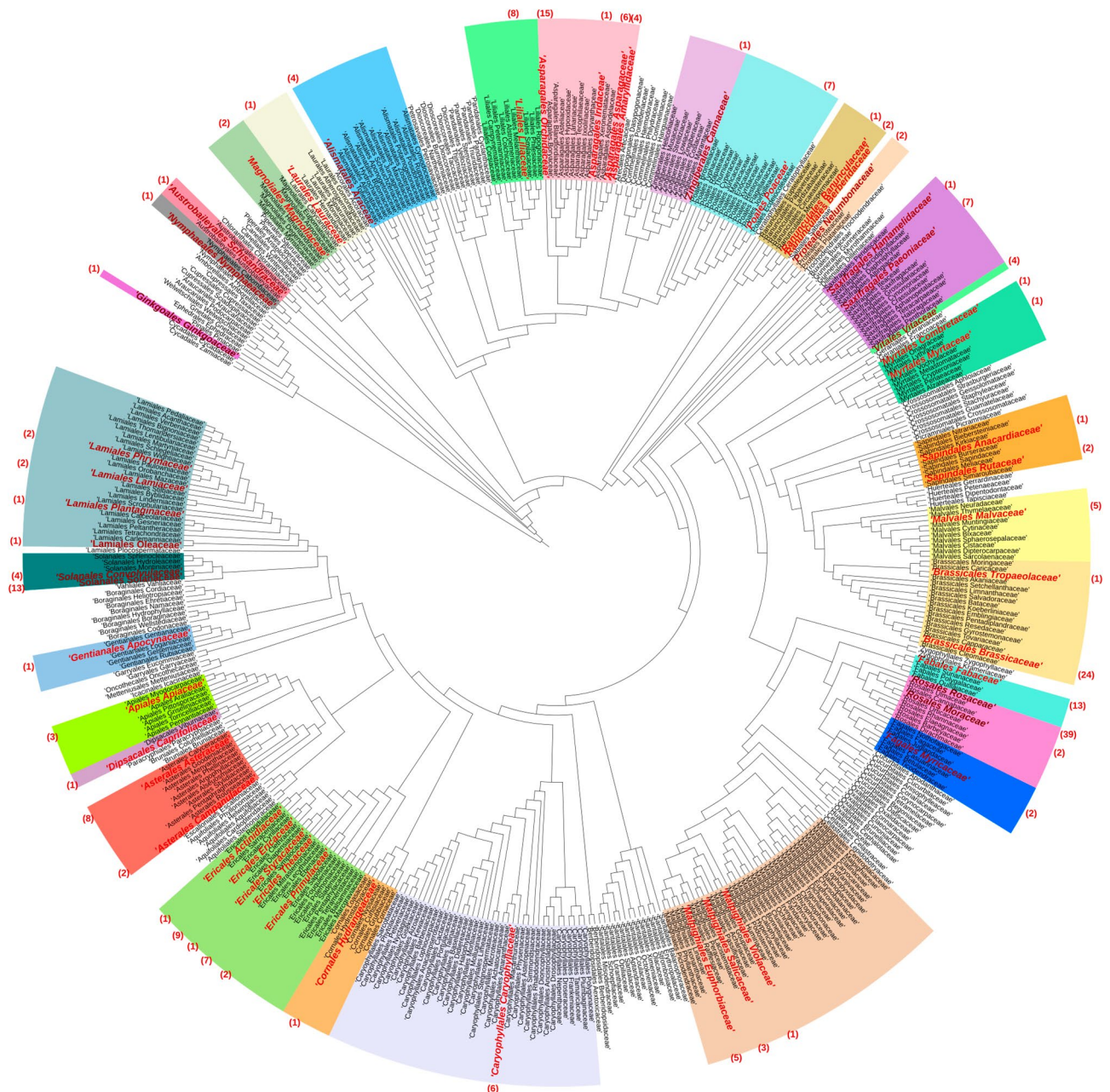


Fig. 2 Phylogenetic tree displaying 428 angiosperm families. Each color range groups the families of an order. Families highlighted in bold red are those encompassed in the literature screening for anthocyanin pigmentation differences. The number of pigmentation difference cases is given in parentheses for each of these families. Tree topology is based on Li et al [95]

This rich diversity in distribution of families and orders is crucial to reveal universal mechanisms explaining pigmentation differences within plant species. Additionally, it highlights the ecological and evolutionary significance of this morphological phenomenon across angiosperms. However, it is noteworthy to clarify that among the different studies the terms “varieties”, “lines”, and “cultivars” were often used interchangeably. In the literature, it was not consistently clarified whether these distinctions arose from horticultural/artificial interventions, or if those differences could be attributed to natural causes.

Genetic hotspots responsible for anthocyanin pigmentation differences

To find out whether mutations in specific anthocyanin biosynthesis genes are predominantly responsible for the loss of anthocyanin pigmentation, reports about pigmentation loss were screened. Based on a total of 235 analyzed cases (see Additional file 1: Table S1), we determined the genes most probable to be responsible for color variation between accessions in each of these species. Four of the included studies did not report one causal gene, but provided the necessary data for a reanalysis (Additional file 2). In the schematic representation (Fig. 3), we defined pigmentation to be the wild type state, while absence of anthocyanin was defined to be the result of a mutation. We identified 13 cases in which upregulated structural genes in pathways competing for substrate with the anthocyanin biosynthesis as the cause of color difference between unpigmented and anthocyanin-pigmented accessions have been reported. Additionally, 58 events of non-activated, down-regulated, non-functional, or lost structural anthocyanin biosynthesis genes were reported in the literature. Moreover, in 147 different cases a transcription factor was proposed to be responsible for differences in pigmentation. Many of these reports named a specific transcription factor. In total, 82 MYBs (activators and repressors), 10 bHLHs, two TTG1 homologs, one bZIP, one WRKY were presented as the causal gene for pigmentation differences. The remaining 49 cases are probably due to the action of multiple transcription factors or caused by TFs that activate the components of the MBW complex. Ten reports presented genes that encode proposed intracellular transporters of anthocyanins as best candidates such as MATE and possibly GST. It was not possible to determine the causal gene in 17 of the analyzed studies.

Genetic factors of anthocyanin loss across plant families

To evaluate whether different genes are predominantly responsible for the loss of anthocyanin pigmentation in different plant lineages, the analysis described for all data sets above was also performed for individual lineages. The genes that have been identified as influential

factors in driving variation were graphically represented along with the families in which they have been reported (Fig. 4). This visualization aims to uncover potential associations between specific genes and their prevalence across different plant lineages.

A noteworthy observation is the prominence of anthocyanin biosynthesis activating MYB transcription factor genes (classified as AnthoMYBact), which have been reported in 75 cases across all plant families. Families with a particularly high prevalence of AnthoMYBact are *Rosaceae*, *Brassicaceae*, *Orchidaceae*, *Liliaceae*, *Solanaceae*, and *Asteraceae*. On the contrary, some genes such as anthocyanin biosynthesis repressing MYBs (AnthoMYBrep), transcription factor bZIP, enzyme CHI, and FlavonolMYB appear in fewer instances, indicating a rare involvement in color variation. By juxtaposing genes associated with anthocyanin loss with the respective families where they have been observed, we aim to discern any family-specific patterns.

Overrepresentation of genetic factors causing anthocyanin loss

The hypothesis that *DFR* might be more often responsible for an anthocyanin loss than *ANS* or other structural genes in the anthocyanin biosynthesis was tested. Among the 235 cases analyzed (see Additional file 1: Table S1), *DFR* exhibited the highest frequency, being reported in 21 cases, while *ANS* was identified in three cases as the gene reported as causal for the color variation. The number of cases reporting *DFR* as the primary factor for pigmentation differences is significantly higher than the number of cases reporting *ANS* (χ^2 test, adjusted p-value=0.0014).

To further examine whether this dominance of *DFR* extends to subsequent anthocyanin genes, we compared *DFR* with *arGST* which also revealed a significant difference (χ^2 test, adjusted p-value=0.0228). Similarly, when examining the relationship between *DFR* and *UGT*, a notable difference was again detected (χ^2 test, adjusted p-value=0.0078). This suggests that *DFR* is the most important target of evolutionary events blocking the anthocyanin pathway through mutations in structural genes.

To investigate if the color variation could be attributed to the substrate competition between the two enzymes *FLS* and *DFR*, we examined the prevalence of cases revealing *FLS* and *DFR* as the primary genes influencing color variation in plant tissue. While the mechanisms underlying *FLS* up-regulation and *DFR* down-regulation are different, the metabolic consequences in terms of relative *FLS* to *DFR* enzymatic activity are similar. A hyperactivation of *FLS* was identified in 10 cases as the factor responsible for anthocyanin pigmentation loss which is significantly lower

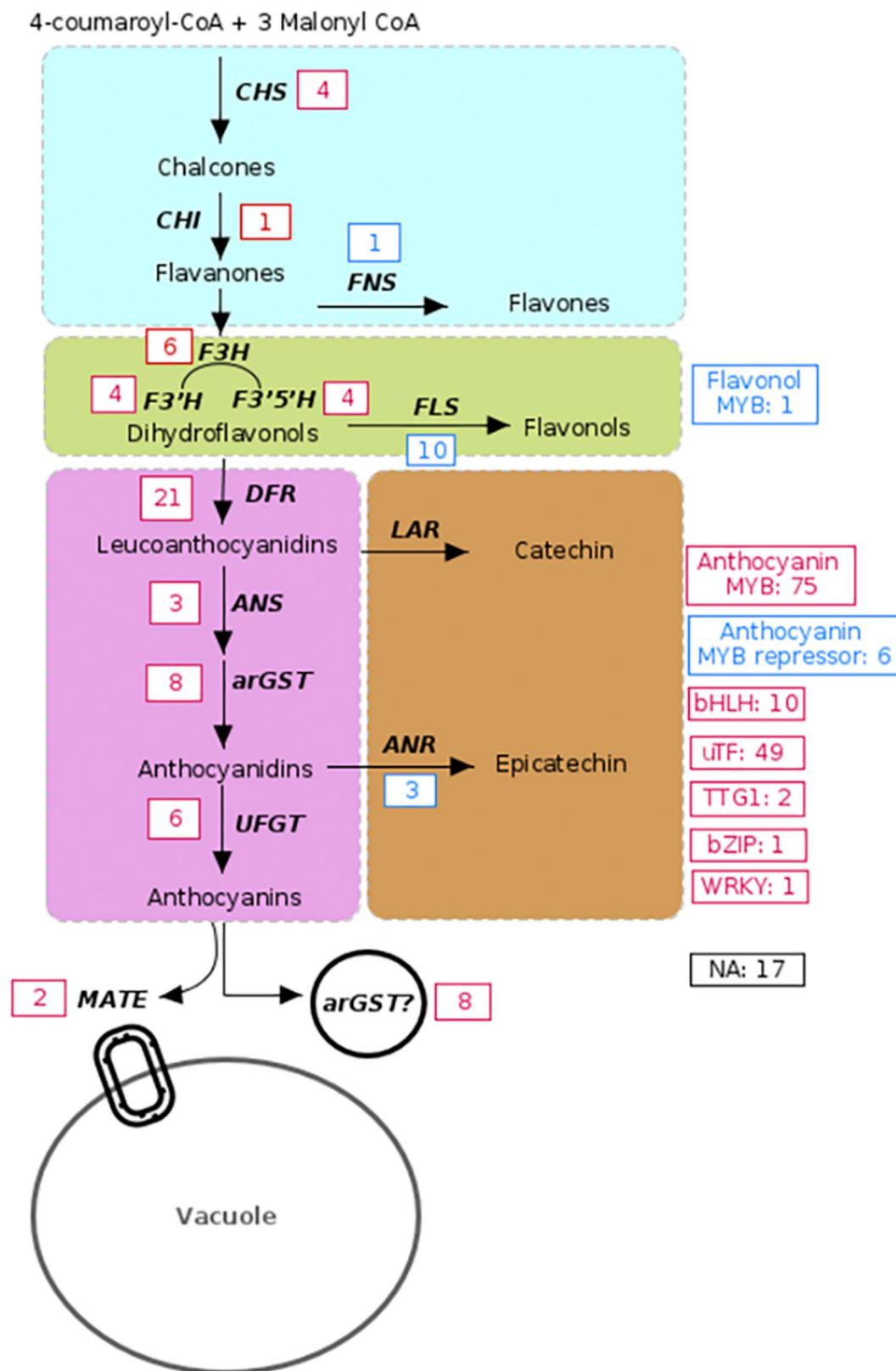


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Schematic representation of the flavonoid biosynthesis pathway with emphasis on the number of cases in which a particular gene was responsible for color difference according to a systematic literature analysis and re-analyses of RNA-Seq data sets. The anthocyanin-pigmented accession is set as reference when determining up- and down-regulation. Blue boxes and red boxes indicate the number of up-regulated and down-regulated/non-functional/lost genes, respectively. Down-regulated genes are placed in one group with genes that lost their function due to mutations in the coding sequence or completely lost genes, because the ultimate function of the gene is lost in any of these cases. Flavonoids were divided into four groups that are indicated by color shading: flavones, flavonols, anthocyanins, and proanthocyanidins. Black bold letters represent the different genes encoding the enzymes and transporters involved in the pathway. *CHS* - chalcone synthase, *CHI* - chalcone isomerase, *F3H* - flavanone 3-hydroxylase, *F3'H* - flavonoid 3'-hydroxylase, *F3'5'H* - flavonoid 3',5'-hydroxylase, *DFR* - dihydroflavonol 4-reductase, *ANS* - anthocyanidin synthase, *UFGT* - UDP-glucose: flavonoid 3-O-glucosyltransferase, *FLS* - flavonol synthase, *LAR* - leucoanthocyanidin reductase, *ANR* - anthocyanidin reductase, *FNS* - flavone synthase, arGST - anthocyanin-related glutathione S-transferase, *MATE* - multidrug and toxin extrusion, anthocyanin MYB - anthocyanin MYB activator, bHLH - basic Helix-Loop-Helix, bZIP - basic leucine zipper, TTTG1- TRANSPARENT TESTA GLABRA1, WRKY- WRKY DNA-binding domain, uTF - unclassified transcription factor, NA - undetermined genetic factor. GST is presented twice due to the recently reported enzymatic function of arGSTs by Eichenberger et al. [38]. It is currently not clear if GST functions only as an enzyme or if it also plays a role in anthocyanin transport

than the number of 21 cases in which a *DFR* down-regulation/loss was responsible (χ^2 test, adjusted p-value=0.048). This suggests a more pronounced influence of *DFR* in the observed color variations, implying a potentially pivotal role in the genetic mechanisms governing anthocyanin production. It further suggests that a down-regulation or silencing of *DFR* is more strongly associated with the white coloration of plant tissues compared to an increased activity in expression of an *FLS* gene and subsequent production of flavonols.

To understand the relative importance of different transcription factors in anthocyanin loss events, the numbers of observed cases were compared as described above for the structural genes. Similar to the comparison between the structural genes, the differences between the TFs reported to be the causal factor of anthocyanin pigmentation differences were analyzed. The chi-square analysis revealed that the frequency of anthocyanin biosynthesis activating MYBs appearing as causal gene of color variation is significantly higher than the presence of other transcription factors such as bHLH, WRKY, TTTG1, bZIP, and others (χ^2 test, adjusted p-value=9.66e-10). Even when compared against the large group of unclassified TFs, we observed that the presence of MYBs is significantly higher (χ^2 test, adjusted p-value=0.0228). In summary, these findings collectively highlight the prevalent role played by the MYB transcription factor family in influencing the observed pigmentation variations.

Discussion

Anthocyanins are one of the main factors responsible for color variation in plant tissues, particularly in flowers. The variation in floral coloration can occur as a result of plant adaptation to different biotic and abiotic conditions, but interactions with pollinators might be the most important function of anthocyanins in flowers [105, 106]. In some cases, flower coloration changes due to visitation by insects [107]. Flavonoids are also known to protect against UV radiation [108], drought [109], and cold stress [26, 110]. Previous studies have reported numerous genes responsible for anthocyanin pigmentation

differences within a plant species (Additional file 1: Table S1). We present an aggregated analysis of the most likely candidate genes responsible for color variation. This analysis harnessed public RNA-Seq data sets that enabled a direct comparison of anthocyanin biosynthesis gene activity between differently pigmented accessions and the taxonomic family they belonged to. This comparison evaluates whether specific genes are responsible for pigmentation loss in certain lineages. The complex interplay between specific genes and their role in shaping plant pigmentation has been a subject of investigation in numerous studies [90, 100, 111–114] and has been experimentally tested in families such as *Solanaceae* [113], as well as in specific genera like *Ipomoea* [73, 75], *Iris* [115], *Antirrhinum* [116], and *Petunia* [93].

DFR is the block 'hotspot' in the anthocyanin branch

DFR activates the conversion of dihydroflavonols to leucoanthocyanidins, which is often considered as the first committed step of the anthocyanin biosynthesis. Through a comprehensive literature survey, we revealed that *DFR* is more often harboring a disruptive variation that results in a block of the anthocyanin accumulation in colorless varieties than any downstream gene in the anthocyanin biosynthesis. A transition from red/purple to white/cream flower color would require some kind of blockage in the anthocyanin production, which probably occurs upstream of anthocyanidin formation [117]. Leucoanthocyanidins can be catalyzed to form two different products, catechins via *LAR* and anthocyanidins via *ANS*. If the anthocyanin biosynthesis is blocked at the *ANS* step, the product to be formed would be catechins. This redirection in the metabolic flow would form proanthocyanidins instead of anthocyanins and could ultimately result in brownish pigmentation due to polymerized and oxidized proanthocyanidins [118].

A study conducted by Rausher et al. [119] proposed that the evolutionary rate of enzymes depends on their location in a pathway with early genes showing a slower evolution rate, but this has been contradicted [120–122]. According to the Arabidopsis Information Resource (TAIR), the length of the coding sequence (CDS) of *DFR*

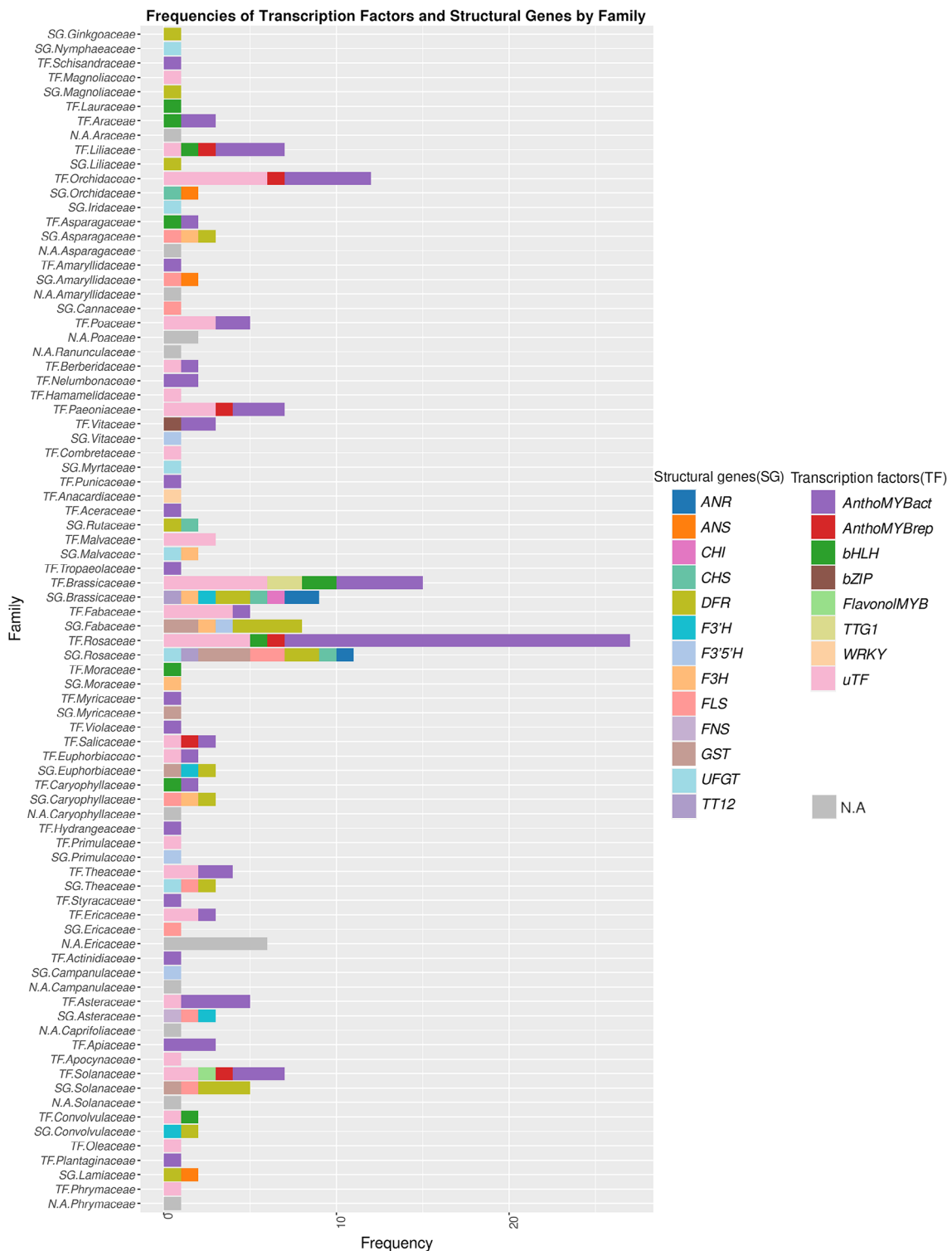


Fig. 4 Number of cases each gene is implicated in anthocyanin differences reported in the literature (Additional file 1: Table S1) resolved by family. ANR – anthocyanidin reductase, ANS – anthocyanidin synthase, AnthoMYBact – anthocyanin MYB activator, AnthoMYBrep – anthocyanin MYB repressor, bHLH – basic Helix-Loop-Helix, bZIP – basic leucine zipper, CHS - chalcone synthase, CHI – chalcone isomerase, DFR – dihydroflavonol 4-reductase, F3'5'H – flavonoid 3',5'-hydroxylase, F3'H – flavonoid 3'-hydroxylase, F3H – flavanone 3-hydroxylase, FLS – flavonol synthase, FNS – flavone synthase, GST (arGST) – anthocyanin-related glutathione S-transferase, uTF – unclassified Transcription Factor, TT12 – TRANSPARENT TESTA12, TTG1 – TRANSPARENT TESTA GLABRA1, UFGT – UDP-glucose: flavonoid 3-O-glucosyltransferase, WRKY – WRKY DNA-binding domain, N.A – undetermined genetic factor

is 1149 bp (accession: NM_123645) and of *ANS* is 1071 bp (accession: NM_118417). To the best of our knowledge and based on the low length difference between the *DFR* and *ANS* coding sequences, there is no evidence that the occurrence of a mutation in *DFR* is substantially more likely than a mutation in *ANS*. While the mutation rate in both genes is probably equal, the rate of mutation fixation might be very different. According to theories of metabolic regulation, it is evolutionary beneficial to have blocks at the first committed step of a branch in a biosynthesis pathway in order to avoid a waste of energy and resources by pushing substrate into a blocked pathway [123, 124]. This could explain why *DFR* and not *ANS*, *arGST*, or *UFGT* appears frequently in analyzed cases of intraspecific anthocyanin pigmentation differences.

Cross talk and substrate competition: anthocyanins vs. flavonols

Plants have multiple mechanisms to regulate their metabolism in response to environmental conditions and availability of resources. Substrate competition is among the factors determining the color variation observed in plant tissues [19]. Metabolically, this can occur when two enzymes or transporters compete for the same or very similar substrates [123, 125]. *DFR* and *FLS* are both catalyzing reactions that utilize dihydroflavonols, but lead to different products. While *DFR* generates colorful anthocyanins, *FLS* produces colorless flavonols. There are three different types of dihydroflavonols namely dihydrokaempferol, dihydroquercetin, and dihydromyricetin that differ in their hydroxylation pattern. Different isoforms of *DFR* and *FLS* have preferences for specific hydroxylation patterns which could be a mechanism to avoid direct substrate competition [125]. The relative activities of F3H, F3'H, and F3'5'H determine the intracellular levels of the three dihydroflavonols. Our analyses revealed that variation associated with *DFR* is more often responsible for a color change than variation associated with *FLS*. In total, 21 cases revealed that the low expression of *DFR* is responsible for the color contrast between unpigmented tissues and those that show anthocyanin pigmentation. Only ten cases showed an increased *FLS* activity as the cause of pigmentation loss. A high expression of *FLS* leads to an accumulation of colorless flavonols instead of colorful anthocyanins as reported previously in several plant species [126]. A recent study identified a flavonol biosynthesis regulating MYB as the most frequently affected gene in pigmentation pattern change [122]. Loss of expression or loss of a gene function can be the consequence of many different mutations and thus be more likely to happen than a gain-of-function mutation. It is also feasible that some researchers only investigated the classical anthocyanin biosynthesis genes when looking for a molecular mechanism to explain the anthocyanin pigmentation difference thus leading to an observation bias

concerning the responsible genes. However, this is unlikely to explain the strong difference between hotspots like *DFR* and MYB and other genes. Once the anthocyanin biosynthesis is disrupted, selection against additional mutations in the anthocyanin biosynthesis genes might be weak or completely absent. This could result in the accumulation of secondary mutations. A number of additional mutations in the anthocyanin biosynthesis would increase the chances that at least one of them is picked up by researchers looking at anthocyanin biosynthesis genes. Performing future analyses by inspecting a more comprehensive gene set could make the identification of causal genes in color difference studies more accurate.

Transcription factor variations appear frequently as block to anthocyanin accumulation

It is well known that the transcriptional activity of structural anthocyanin biosynthesis genes is regulated by a ternary complex consisting of a MYB, a bHLH and a WD40 protein (MBW complex). The anthocyanin biosynthesis is even considered a model system for transcriptional control in eukaryotes. Previous studies identified transcriptional activation of different R2R3-MYBs [127–130] and bHLHs [131–133] as the cause for the increase in anthocyanin levels. Our results showed that transcription factors were three times more often reported as causal factors of color differences than structural genes. While we cannot rule out the possibility that structural genes can also accumulate variation prior to the reduction in transcription, this observation is in line with a previous study that observed faster evolutionary changes in transcription factors than in structural genes [124]. Similarly, Wheeler et al. [122], showed that transcription factors, particularly MYBs, presented lower levels of gene expression with higher molecular evolutionary rate compared to their targeted structural genes, and suggested a negative correlation between evolution rate and gene expression in the *Petuniae* tribe [122]. This premise commonly known as the E-R anticorrelation, has been widely studied across different organisms, including yeast [134, 135], *Arabidopsis* [136], *Brassica* [137], Barley [138], *Arachis* [139], and *Drosophila* [140]. However, the hypotheses explaining this model are still a topic of debate.

The loss of a transcription factor can switch off an entire biosynthesis pathway, while genes involved in this pathway could still be activated by other transcription factors to harness their activity in a different metabolic context. For example, *DFR* and *ANS*, two important anthocyanin biosynthesis genes, are also required for the biosynthesis of proanthocyanidins. A selective loss of anthocyanins and maintenance of the proanthocyanidin biosynthesis can not be caused by the loss of *DFR* or *ANS*. A well known example for such a scenario is the

conservation of *DFR* and *ANS* across betalain-pigmented lineages of the *Caryophyllales* [141], which do not accumulate anthocyanins [80, 126].

Studies on evolutionary rates investigating the components of the MBW complex suggested that MYBs would be the most probable component to be lost due to the highest degree of specialization which coincides with a lower pleiotropy [74, 122, 142]. It was observed that insertions/deletions were the most frequent mutation events in MYBs, while amino acid substitutions in the conserved region appeared irrelevant [73]. This led to the hypothesis that amino acid substitutions in transcription factors might not be relevant for the pigment evolution context, while InDels could disrupt the function of the encoded protein [73]. However, a recent investigation suggests a high importance of amino acid substitutions in the R3 interaction domain of MYBs in the loss of anthocyanin pigmentation in betalain-pigmented *Caryophyllales* [141]. These amino acid substitutions alter a highly conserved region that is considered crucial for the interaction of MYB and bHLH protein in the MBW complex which is required for activation of anthocyanin biosynthesis genes [143–145]. The lack of a functional MBW complex is considered as one crucial factor for the loss of anthocyanin pigmentation in the betalain-pigmented *Caryophyllales* [141]. In summary, there is evidence for InDels and amino acid substitutions as mechanisms that can disrupt the function of MYBs involved in the pigmentation biosynthesis regulation.

The frequency of mutations in TFs has been studied in various species. For example, a study of the *A. thaliana* genome revealed that there are more than 2,000 genes that encode for TFs and that these genes were more prone to accumulate mutations than non-TF genes [146]. Another study examined the frequency of point mutations in TF genes in *Escherichia coli* and found that these genes were more vulnerable to harmful mutations that resulted in significant changes in gene expression than non-TF genes [147]. Given that their activity covers a wide range of functions, transcription factors can regulate and affect the expression of structural genes without the need for high gene expression levels [122]. This multifunctionality becomes apparent when examining specific TF families, such as MYBs, which are crucial in the regulation of the anthocyanin biosynthesis. In this context, a recent study by Liang *et al.* discovered a single point mutation in the 5'-UTR of *PELAN*, an anthocyanin-activating R2R3-MYB, as causal mutation for the loss of pigmentation in *Mimulus parishii* [148]. The results of their experiments concluded that despite the similar transcript abundance of *PELAN* in both strong pigmented and low pigmented cultivars, the difference in phenotype was due to a mistranslation of the *PELAN* mRNA [148].

The number of TFs required to regulate a specific enzyme-encoding gene is a complex and dynamic

process that is influenced by various factors [149], including the complexity of the regulatory region [150], the stage of development or cell type [151], and environmental factors [152]. For example, only about ten *R2R3-MYBs* are known to bind specific DNA motifs related to the regulation of the flavonoid biosynthesis pathway in *A. thaliana* [57, 153]. In a review study, Feller *et al.* [70], performed a comparative analysis of the transcription factor families MYB and bHLH. They concluded that the reason why the bHLH family contains one of the largest numbers of transcription factors in plants is due to their functional diversification [70]. Multiple bHLH proteins contain a similar ligand-binding domain targeting different enzyme encoding genes [154]. A greater proportion of MYBs were recognized to be responsible for the regulation of flavonoid biosynthesis genes in comparison to bHLHs [154]. Our results align with this observation, because MYBs were reported as the causal gene of color differences in 75 cases, while bHLHs were only reported as a crucial factor in ten cases. If bHLHs are more often involved in multiple processes, their loss would be more detrimental which makes it less likely to occur. Similar explanation can be inferred from contrasting the reported cases involving MYBs with those involving other TF, such as, TTG1, WRKY, and bZIP. Given that most classified transcription factors are MYBs, it is expected that most of the cases with unidentified transcription factors would actually be due to anthocyanin biosynthesis activating MYBs.

Conclusion

Our systematic literature screening supported the assumption that variations in transcription factors are the most frequently observed blocks in the anthocyanin accumulation in cases of intraspecies pigmentation differences. Specifically, the MYB components of the MBW complex exhibited dominance in influencing anthocyanin accumulation variations among differently pigmented accessions. The degree of transcription factor specialization for a certain pathway seems to determine the frequency of their implication in color differences with more pleiotropic TFs like bHLH and TTG1 having a lower relevance. According to our results, MYBs are most often responsible for the difference in anthocyanin content, followed by bHLH, and other TFs. When structural genes appeared to be responsible for the absence of anthocyanins, this was most often a lack of *DFR* activity. An increased activation of the flavonol biosynthesis as a pathway competing for substrate with the anthocyanin biosynthesis was seen in rare cases. Therefore, our findings highlight the pivotal role of transcription factors, particularly MYBs, in determining anthocyanin content differences within species.

Methods

Extensive literature screening

An extensive record identification was performed in electronic databases (PubMed, Google Scholar, JSTOR) using specific screening keywords: “flower pigmentation differences”, “leaf pigmentation”, “color difference”, and “anthocyanin loss”. A total of 230 full-text articles were included for an eligibility assessment between December 2021 to October 2023. Accessible articles were considered if the genetic basis of pigmentation was investigated in the respective study (Additional file 1: Table S1). The evidence for causal genes reported in the literature differs between studies. We classified the presented evidence into the following main categories: ‘knockout mutant’, ‘in vitro characterization’, ‘coexpression patterns’, ‘metabolic exploration’, among others (Additional file 1: Table S1). Additionally, the respective family and order of each investigated species were collected along with the names of accessions, varieties, lines, or cultivars involved in the study. Most studies compared accessions of the same species that differed in anthocyanin pigmentation thus we are predominantly exploring intraspecific mechanisms of anthocyanin loss. While studies were classified as intraspecific or interspecific, we refrained from separate analyses due to a low sample size. As the classification of plants into categories like accessions, subspecies, and closely related species might leave some room for discussion, we considered all these studies.

Data sources

From the 230 different articles, four studies were selected for in-depth transcriptome re-analysis. The four studies generated RNA-Seq data sets for the analysis of genetic factors underlying differences in pigmentation. The analyzed species were *Michelia maudiae* [155], *Rhododendron obtusum* [156], *Trifolium repens* [157], and *Hosta plantaginea* [158] (Additional file 2). The selection of each dataset was based on the following criteria: (I) paired-end RNA-Seq data, (II) study has biological replicates, and (III) the authors have not identified the specific gene responsible for the color difference. The RNA-Seq datasets of those four plant species were retrieved from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) (Additional file 1: Table S2) using fastq-dump [159].

De novo transcriptome assembly

Transcriptomic data sets of four plant species were re-analysed to identify a candidate gene that could explain the absence of anthocyanins in one accession of each of these species. The generation of *de novo* transcriptome assemblies was necessary, because no transcriptome or genome sequences of these species were

publicly available. Trimmomatic v0.39 [160] was used to remove adapter sequences, to eliminate leading and trailing low-quality reads with quality below 3 (LEADING:3, TRAILING:3), and to drop reads shorter than 36 nt (MINLEN:36). The IDs of all remaining reads were modified by a customized Python script [161] to enable the following assembly with Trinity [162]. Trinity v2.4.0 [162] was applied for *de novo* transcriptome assembly using the previous cleaned reads as input. Trinity was run with a k-mer length of 25. In order to validate the quality of the transcriptome assemblies, a summary of the assembly statistics was generated for each species. The assembly statistics were computed using a previously developed Python script [163] (Additional file 1: Table S3).

After completion of the assembly process, kallisto v0.44 [164] was run to quantify the abundances of transcripts based on all available RNA-Seq data of the respective species. Similarly, a principal components analysis (PCA) for every dataset was performed based on their transcriptomic profiles to compare the similarity between samples and to identify any outliers (Additional file 3: Figures S1-S4). PCA was performed using R v.4.1.3 [165] with the package ggplot2 v.3.4.0 [166] to inspect the variation within the data set (Additional file 3: Figures S1-S4).

Identification of candidate genes

In order to facilitate the identification of candidate genes, encoded peptide sequences were inferred from the transcriptome assembly using a previously established approach [167] that combines Transdecoder [168], ORFfinder [169], and ORFpredictor [170]. Knowledge-based Identification of Pathway Enzymes (KIPes3) v0.34 [171, 172] was applied to identify the structural genes involved in the flavonoid biosynthesis. Flavonoid biosynthesis regulating R2R3-MYB genes were identified via MYB_annotator v0.3 [173]. A previously described BLAST-based Python script [174] was deployed for the identification of additional candidate genes using bHLH, WD40, and WRKY genes associated with the flavonoid biosynthesis as baits [161]. A complete list of the selected candidate genes can be found in Additional file 1: Tables S4-S7.

Phylogenetic trees were constructed to identify all isoforms that belong to the same gene. Isoforms of the same gene may differ by the presence or absence of exons, but they should not exhibit more sequence differences than those accounted for by sequencing errors. In a phylogenetic context, transcript isoforms should form a distinct monophyletic group that can be replaced by one representative sequence. Phylogenetic trees were constructed with FastTree v.2.1.11 [175] based on a MAFFT v7.475 alignment

of polypeptide sequences using default parameters. Additional trees for comparison and additional support were constructed using IQ-TREE v.1.6.12 [176] using default parameters and RAxML v.8.2.12 [177] with PROTGAMMA+LG+F. Phylogenetic trees for the transcription factor families MYB, bHLH, TTG1, and WRKY were constructed to assess orthologues relationships (Additional file 3: Figures S8-S17). Sets of outgroup sequences were compiled based on reports in the literature in order to have a backbone of functionally characterized sequences for each tree. These sequences have been associated with the flavonoid biosynthesis in previous studies and were taken from plant species closely related to those explored with transcriptome assemblies.

Taxonomic distribution of analyzed species

A plastid-based phylogenomic tree was used to study the distribution of the variations-related cases across all flowering plant families. The backbone tree was taken from Li et al [178] and modified with iTOL v.6.8.1 [179] to highlight the represented orders and families in our dataset.

Gene expression analyses

Transcriptome assemblies usually generate a huge number of alternative transcript isoforms per gene. The initial sequences of the transcriptome assembly were used as reference for the quantification, but the transcript abundance ('gene expression') values were summarized per gene. Gene expression information of the entire monophyletic group was mapped to this representative transcript during the generation of heatmaps. Heatmaps displaying the candidate genes and their respective abundance as transcripts per million (TPM) were constructed using the R packages ComplexHeatmap v.2.10.0 [180], circlize 0.4.15 [181], and dplyr 1.1.0 [182]. Genes with an adjusted p-value < 0.05 and absolute log₂ fold-change > 1 were considered as differentially expressed. The script used for the heatmap construction is available in our GitHub repository [161]. A workflow of the methods used in the transcriptome analysis is described in Fig. 5.

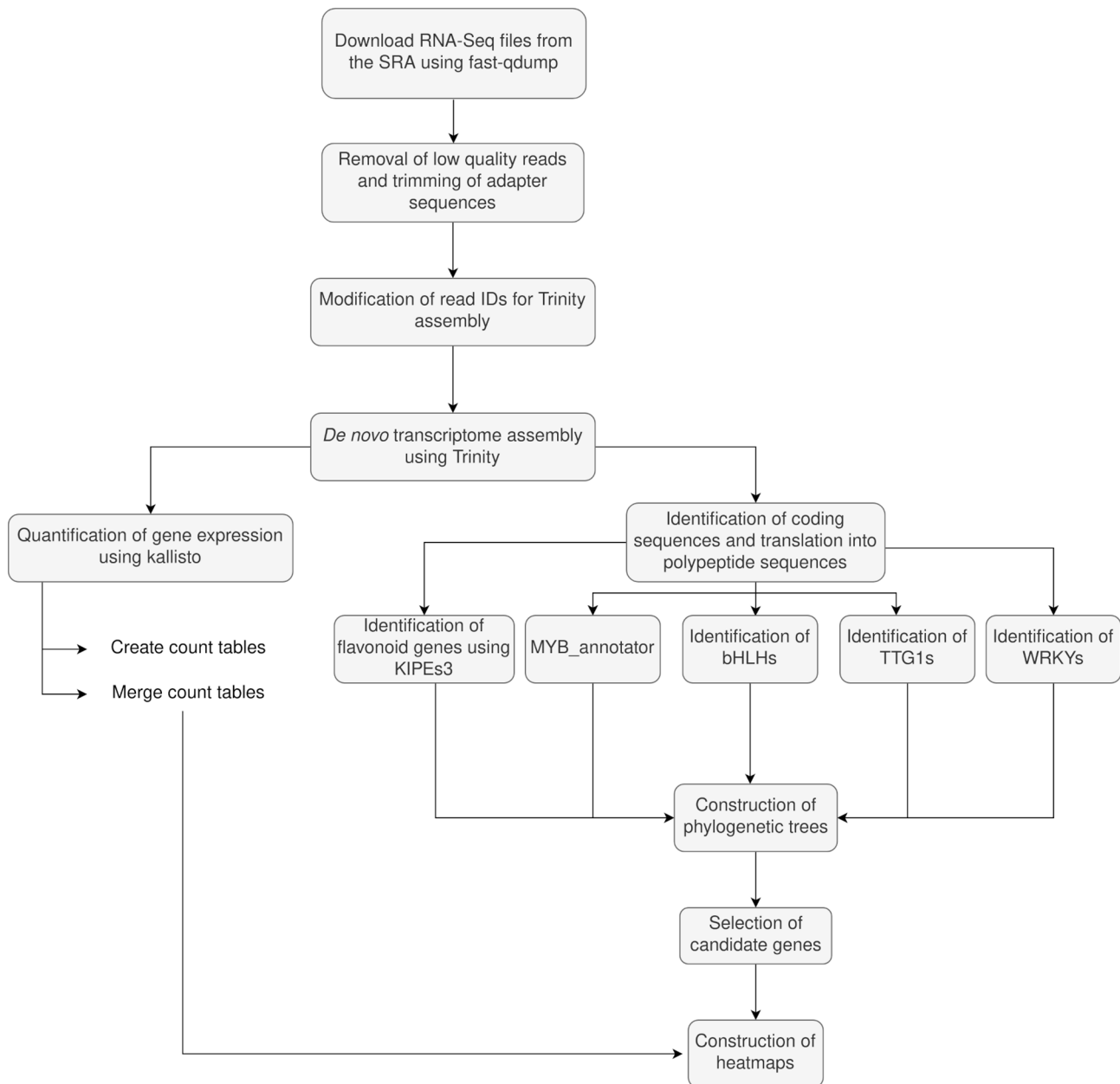


Fig. 5 Flowchart representation of the methodology followed in the comparative transcriptional analysis. SRA – Sequence Read Archive, KIPes – Knowledge-based Identification of Pathway Enzymes, MYB - Myeloblastosis, bHLH – basic helix-loop-helix, TTG1 – TRANSPARENT TESTA GLABRA1, WRKY – WRKY DNA-binding domain

Abbreviations

PAL	phenylalanine ammonia-lyase
C4H	cinnamic acid 4-hydroxylase
4CL	4-coumarate-CoA ligase
CHS	chalcone synthase
CHI	chalcone isomerase
F3H	flavanone 3-hydroxylase
F3'H	flavonoid 3'-hydroxylase
F3'5'H	flavonoid 3';5'-hydroxylase
DFR	dihydroflavonol 4-reductase
ANS	anthocyanidin synthase
UFGT	UDP-glucose: flavonoid 3-O-glucosyltransferase
FLS	flavonol synthase
ANR	anthocyanidin reductase
LDOX	leucoanthocyanidin dioxygenase

FNS	flavone synthase
ER	Endoplasmic Reticulum
MATE	Multidrug and Toxin Extrusion Transporter
GST	Glutathione S-transferase
ABCC	ATP-binding cassette
SRA	Sequence Read Archive
KIPes	Knowledge-based Identification of Pathway Enzymes
MYB	Myeloblastosis
bHLH	basic Helix-Loop-Helix
TTG1	TRANSPARENT TESTA GLABRA1
WRKY	WRKY DNA-binding domain

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05316-w>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A). We are grateful to de.NBI and the Center for Biotechnology (CeBiTec) at Bielefeld University for providing us with the resources and setting to conduct the necessary computational analyses. We are also thankful to the members of the Plant Biotechnology and Bioinformatics group at TU Braunschweig for their helpful discussions and valuable comments during the manuscript revision process. We acknowledge support by the Open Access Publication Funds of Technische Universität Braunschweig.

Author contributions

BP conceptualized the research project. MFMR carried out the analyses and prepared figures. MFMR and BP wrote the manuscript. Both authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work received no external funding.

Open Access funding enabled and organized by Projekt DEAL.

Data availability

All the transcriptome data used in this study was obtained from the NCBI Sequence Read Archive (SRA) under accession numbers PRJNA504531, PRJNA542483, PRJNA700000, and PRJNA393638 (<http://www.ncbi.nlm.nih.gov/sra>). The data charts supporting the results and conclusions are included in the article and additional files. All the assemblies and scripts used in reanalyzed transcriptome analysis have been deposited in our GitHub repository (<https://github.com/bpucker/codi>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Statistical analysis

The observed distribution of frequency values in relation to the specific genes hotspots of color differences in the 230 studies, were evaluated using a 2-factor Chi-Square test. P values were adjusted for multiple comparisons using the FDR correction method for multiple testing. All the tests were applied using the statistical software R v.4.1.3 [165].

Received: 15 November 2023 / Accepted: 20 June 2024

Published online: 03 July 2024

References

1. Alkema J, Seager SL. The chemical pigments of plants. *J Chem Educ*. 1982;59(3):183.
2. Davies KM, Albert NW, Schwinn KE, Davies KM, Albert NW, Schwinn KE. From landing lights to mimicry: the molecular regulation of flower

- colouration and mechanisms for pigmentation patterning. *Funct Plant Biol*. 2012;39(8):619–38.
3. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J*. 2008;54(4):733–49.
4. Grotewold E. The Genetics and Biochemistry of Floral pigments. *Annu Rev Plant Biol*. 2006;57(1):761–80.
5. Ruiz-Sola MA, Rodríguez-Concepción M. Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. *Arab Book [Internet]*. 2012 Jan [cited 2023 Nov 7];2012(10). <https://bioone.org/journals/the-arabidopsis-book/volume-2012/issue-10/tab.0158/Carotenoid-Biosynthesis-in-Arabidopsis-Ahttps://doi.org/10.1199/tab.0158.full>.
6. Burns KC, Dalen JL. Foliage color contrasts and adaptive fruit color variation in a bird-dispersed plant community. *Oikos*. 2002;96(3):463–9.
7. Gould KS. Nature's Swiss Army Knife: the diverse protective roles of anthocyanins in leaves. *J Biomed Biotechnol*. 2004;2004(5):314–20.
8. Chutipajit S, Cha-um S, Sompornpailin K. High contents of proline and anthocyanin increase protective response to salinity in *Oryza sativa* L. spp. *indica*. *Aust J Crop Sci*. 2011;5(10):1191–8.
9. Hernández I, Alegre L, Munné-Bosch S. Drought-induced changes in flavonoids and other low molecular weight antioxidants in *Cistus clusii* grown under Mediterranean field conditions. *Tree Physiol*. 2004;24(11):1303–11.
10. Lois R. Accumulation of UV-absorbing flavonoids induced by UV-B radiation in *Arabidopsis thaliana* L.: I. mechanisms of UV-resistance in Arabidopsis. *Planta*. 1994;194(4):498–503.
11. Kootstra A. Protection from UV-B-induced DNA damage by flavonoids. *Plant Mol Biol*. 1994;26(2):771–4.
12. Samanta A, Das G, Das SK. Roles Flavonoids Plants. 2011;6(1):12–35.
13. Abid MA, Wei Y, Meng Z, Wang Y, Ye Y, Wang Y, et al. Increasing floral visitation and hybrid seed production mediated by beauty mark in *Gossypium hirsutum*. *Plant Biotechnol J*. 2022;20(7):1274–84.
14. Whittall JB, Hodges SA. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*. 2007;447(7145):706–9.
15. Thompson WR, Meinwald J, Aneshansley D, Eisner T. Flavonols: pigments responsible for Ultraviolet absorption in Nectar Guide of Flower J. *Science*. 1972;177.
16. Yonekura-Sakakibara K, Tohge T, Niida R, Saito K. Identification of a flavonol 7-O-Rhamnosyltransferase gene determining Flavonoid Pattern in Arabidopsis by Transcriptome Coexpression Analysis and Reverse Genetics. *J Biol Chem*. 2007;282(20):32–41.
17. Tohge T, de Souza LP, Fernie AR. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J Exp Bot*. 2017;68(15):4013–28.
18. Yonekura-Sakakibara K, Nakayama T, Yamazaki M, Saito K. Modification and Stabilization of Anthocyanins. Winefield C, Davies K, Gould K, editors. *Anthocyanins Biosynth Funct Appl*. 2009;169–90.
19. Davies KM, Schwinn KE, Derolles SC, Manson DG, Lewis DH, Bloor SJ, et al. Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. *Euphytica*. 2003;131(3):259–68.
20. Mazza G, Brouillard R. The mechanism of co-pigmentation of anthocyanins in aqueous solutions. *Phytochemistry*. 1990;29(4):1097–102.
21. Yabuya T, Nakamura M, Iwashina T, Yamaguchi M, Takehara T. Anthocyanin-flavone copigmentation in bluish purple flowers of Japanese garden iris (*Iris ensata* Thunb). *Euphytica*. 1997;98(3):163–7.
22. Sendri N, Singh S, Sharma B, Purohit R, Bhandari P. Effect of co-pigments on anthocyanins of Rhododendron arboreum and insights into interaction mechanism. *Food Chem*. 2023;426:136571.
23. Horbowicz M, Kosson R, Grzesiuk A, Dębski H. Anthocyanins of fruits and vegetables - their occurrence, analysis and role in Human Nutrition. *J Fruit Ornament Plant Res*. 2008;68(1):5–22.
24. Dick CA, Buenrostro J, Butler T, Carlson ML, Kliebenstein DJ, Whittall JB. Arctic Mustard Flower Color Polymorphism controlled by Petal-Specific Downregulation at the threshold of the anthocyanin Biosynthetic Pathway. *PLoS ONE*. 2011;6(4):e18230.
25. Ghissing U, Goswami A, Mitra A. Temporal accumulation of pigments during colour transformation from white to red in *Combretum indicum* (L.) DeFilippis (syn. *Quisqualis indica* L.) flowers. *Nat Prod Res*. 2021;37(3):529–33.
26. Dai Y, Zhang L, Sun X, Li F, Zhang S, Zhang H, et al. Transcriptome analysis reveals anthocyanin regulation in Chinese cabbage (*Brassica rapa* L.) at low temperatures. *Sci Rep*. 2022;12(1):6308.
27. Koes RE, Quattrocchio F, Mol JNM. The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays*. 1994;16(2):123–32.

28. Schütz K, Kammerer DR, Carle R, Schieber A. Characterization of phenolic acids and flavonoids in dandelion (*Taraxacum officinale* WEB. Ex WIGG.) Root and herb by high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom RCM*. 2005;19(2):179–86.
29. Zhang Z, CHEN SM, LI CH Y-M, DENG X-R, ZHU W. SUN4,5, L.-S. WANG2, F.-D. CHEN1* and Z. ZHANG. The identification of flavonoids and the expression of genes of anthocyanin biosynthesis in the chrysanthemum flowers. *Biol Plant*. 2012;56(3):458–64.
30. Wang Y, Zhou LJ, Wang Y, Liu S, Geng Z, Song A, et al. Functional identification of a flavone synthase and a flavonol synthase genes affecting flower color formation in *Chrysanthemum morifolium*. *Plant Physiol Biochem*. 2021;166:1109–20.
31. Constabel CP. Molecular controls of Proanthocyanidin Synthesis and structure: prospects for Genetic Engineering in Crop plants. *J Agric Food Chem*. 2018;66(38):9882–8.
32. Dixon RA, Xie D, Sharma SB. Proanthocyanidins – a final frontier in flavonoid research? *New Phytol*. 2005;165(1):9–28.
33. Leung J, Fenton TW, Mueller MM, Clandinin OR. Condensed tannins of Rape-seed Meal. *J Food Sci*. 1979;44(5):1313–7.
34. Park KI, Ishikawa N, Morita Y, Choi JD, Hoshino A, Iida S. A bHLH regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation: epidermal traits controlled by bHLH2. *Plant J*. 2007;49(4):641–54.
35. Winkel-Shirley B. Flavonoid Biosynthesis. A colorful model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiol*. 2001;126(2):485–93.
36. Mueller LA, Goodman CD, Silady RA, Walbot V. AN9, a Petunia glutathione S-Transferase required for anthocyanin sequestration, is a flavonoid-binding Protein1. *Plant Physiol*. 2000;123(4):1561–70.
37. Kitamura S, Shikazono N, Tanaka A. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant J*. 2004;37(1):104–14.
38. Eichenberger M, Schwander T, Hüppi S, Kreuzer J, Mittl PRE, Peccati F et al. The catalytic role of glutathione transferases in heterologous anthocyanin biosynthesis. *Nat Catal*. 2023;1–12.
39. Grotewold E, Davies K. Trafficking and sequestration of anthocyanins. *Nat Prod Commun*. 2008;3(8).
40. Kitamura S, Oono Y, Narumi I. Arabidopsis pab1, a mutant with reduced anthocyanins in immature seeds from banyuls, harbors a mutation in the MATE transporter FFT. *Plant Mol Biol*. 2016;90(1):7–18.
41. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci*. 2015;20(9):576–85.
42. Pucker B, Selmar D. Biochemistry and Molecular Basis of Intracellular Flavonoid Transport in plants. *Plants*. 2022;11(7):963.
43. Grotewold E, Chamberlin M, Snook M, Siame B, Butler L, Swenson J, et al. Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. *Plant Cell*. 1998;10(5):721–40.
44. Pourcel L, Irani NG, Lu Y, Riedl K, Schwartz S, Grotewold E. The formation of Anthocyanic Vacuolar inclusions in Arabidopsis thaliana and implications for the sequestration of anthocyanin pigments. *Mol Plant*. 2010;3(1):78–90.
45. Gomez C, Conejero G, Torregrosa L, Cheynier V, Terrier N, Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J*. 2011;67(6):960–70.
46. Mackon E, Ma Y, Jeazet Dongho Epse Mackon GC, Li Q, Zhou Q, Liu P. Subcellular localization and Vesicular Structures of Anthocyanin Pigmentation by Fluorescence Imaging of Black Rice (*Oryza sativa* L.) Stigma Protoplast. *Plants*. 2021;10(4):685.
47. Poustka F, Irani NG, Feller A, Lu Y, Pourcel L, Frame K, et al. A trafficking pathway for anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein-sorting Route in Arabidopsis and contributes to the formation of Vacuolar inclusions. *Plant Physiol*. 2007;145(4):1323–35.
48. Zhao J, Dixon RA. The 'ins' and 'outs' of flavonoid transport. *Trends Plant Sci*. 2010;15(2):72–80.
49. Marrs KA, Alfenito MR, Lloyd AM. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. 1995;375.
50. Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, et al. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell*. 1998;10(7):1135–49.
51. Lu Z, Cao H, Pan L, Niu L, Wei B, Cui G, et al. Two loss-of-function alleles of the glutathione S-transferase (GST) gene cause anthocyanin deficiency in flower and fruit skin of peach (*Prunus persica*). *Plant J*. 2021;107(5):1320–31.
52. Kaur S, Sharma N, Kapoor P, Chunduri V, Pandey AK, Garg M. Spotlight on the overlapping routes and partners for anthocyanin transport in plants. *Physiol Plant*. 2021;171(4):868–81.
53. Debeaujon I, Peeters AJM, Léon-Kloosterziel KM, Koornneef M. The TRANSPARENT TESTA12 gene of Arabidopsis encodes a Multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed Coat Endothelium. *Plant Cell*. 2001;13(4):853–71.
54. Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul JM, et al. The Arabidopsis MATE Transporter TT12 acts as a Vacuolar Flavonoid/H⁺-Antiporter active in Proanthocyanidin-Accumulating cells of the seed Coat. *Plant Cell*. 2007;19(6):2023–38.
55. Mehrtens F, Kranz H, Bednarek P, Weisshaar B. The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol*. 2005;138(2):1083–96.
56. Stracke R, Ishihara H, Huelop G, Barsch A, Mehrtens F, Niehaus K, et al. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J*. 2007;50(4):660–77.
57. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. *Trends Plant Sci*. 2010;15(10):573–81.
58. Ramsay NA, Glover BJ. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci*. 2005;10(2):63–70.
59. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. The TT8 gene encodes a Basic Helix-Loop-Helix domain protein required for expression of DFR and BAN genes in Arabidopsis Siliques. *Plant Cell*. 2000;12(10):1863–78.
60. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J*. 2004;39(3):366–80.
61. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J*. 2008;53(5):814–27.
62. Zhang B, Hülkamp M. Evolutionary analysis of MBW function by Phenotypic Rescue in *Arabidopsis thaliana*. *Front Plant Sci*. 2019;10.
63. Busche M, Pucker B, Weisshaar B, Stracke R, Three. R2R3-MYB transcription factors from banana (*Musa acuminata*) activate structural anthocyanin biosynthesis genes as part of an MBW complex. *BMC Res Notes*. 2023;16(1):103.
64. Jiang L, Yue M, Liu Y, Zhang N, Lin Y, Zhang Y, et al. A novel R2R3-MYB transcription factor FaMYB5 positively regulates anthocyanin and proanthocyanidin biosynthesis in cultivated strawberries (*Fragaria × ananassa*). *Plant Biotechnol J*. 2023;21(6):1140–58.
65. Symonds VV, Hatlestad G, Lloyd AM. Natural allelic variation defines a role for ATM1: Trichome Cell Fate determination. *PLOS Genet*. 2011;7(6).
66. Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*. 2000;12(12):2383–94.
67. Lafferty DJ, Espley RV, Deng CH, Günther CS, Plunkett B, Turner JL, et al. Hierarchical regulation of MYBPA1 by anthocyanin- and proanthocyanidin-related MYB proteins is conserved in Vaccinium species. *J Exp Bot*. 2022;73(5):1344–56.
68. Li Y, Shan X, Gao R, Yang S, Wang S, Gao X, et al. Two Ilif Clade-bHLHs from Freesia Hybrida Play Divergent roles in Flavonoid biosynthesis and trichome formation when ectopically expressed in Arabidopsis. *Sci Rep*. 2016;6(1):30514.
69. Zumajo-Cardona C, Gabrieli F, Anire J, Albertini E, Ezquer I, Colombo L. Evolutionary studies of the bHLH transcription factors belonging to MBW complex: their role in seed development. *Ann Bot*. 2023;132(3):383–400.
70. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J*. 2011;66(1):94–116.
71. Broun P. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. *Curr Opin Plant Biol*. 2005;8(3):272–9.
72. Martin C, Ellis N, Rook F. Do transcription factors play special roles in adaptive variation? *Plant Physiol*. 2010;154(2):506–11.
73. Streisfeld MA, Liu D, Rausher MD. Predictable patterns of constraint among anthocyanin-regulating transcription factors in Ipomoea. *New Phytol*. 2011;191(1):264–74.

74. Streisfeld MA, Rausher MD. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution: mutations during adaptive evolution. *Evolution*. 2011;65(3):629–42.
75. Streisfeld MA, Rausher MD. Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytol*. 2009;183(3):751–63.
76. Smith SD, Rausher MD. Gene loss and parallel evolution contribute to Species Difference in Flower Color. *Mol Biol Evol*. 2011;28(10):2799–810.
77. Almeida J, Carpenter R, Robbins TP, Martin C, Coen ES. Genetic interactions underlying flower color patterns in *Antirrhinum majus*. *Genes Dev*. 1989;3(11):1758–67.
78. Onozaki T, Mato M, Shibata M, Ikeda H. Differences in flower color and pigment composition among white carnation (*Dianthus Caryophyllus* L.) cultivars. *Sci Hortic*. 1999;82(1–2):103–11.
79. Zufall RA. The genetic basis of a Flower Color Polymorphism in the Common Morning Glory (*Ipomoea purpurea*). *J Hered*. 2003;94(6):442–8.
80. Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, Lopez-Nieves S, et al. The evolution of betalain biosynthesis in Caryophyllales. *New Phytol*. 2019;224(1):71–85.
81. Clement JS, Mabry TJ. Pigment evolution in the Caryophyllales: a systematic Overview*. *Bot Acta*. 1996;109(5):360–7.
82. Bate-Smith EC. The phenolic constituents of plants and their taxonomic significance. I. Dicotyledons. *J Linn Soc Lond Bot*. 1962;58(371):95–173.
83. Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, et al. Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J*. 1995;8(5):659–71.
84. Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, et al. Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics*. 2006;7(1):12.
85. Chen Z, Yu L, Liu W, Zhang J, Wang N, Chen X. Research progress of fruit color development in apple (*Malus domestica* Borkh.). *Plant Physiol Biochem*. 2021;162:267–79.
86. Naeem M, Zhao W, Ahmad N, Zhao L. Beyond green and red: unlocking the genetic orchestration of tomato fruit color and pigmentation. *Funct Integr Genomics*. 2023;23(3):243.
87. Shoeva OY, Mock HP, Kukoeva TV, Börner A, Khlestkina EK. Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of *Hordeum vulgare*. *PLoS ONE*. 2016;11(10):e0163782.
88. Jiao F, Zhao L, Wu X, Song Z, Li Y. Metabolome and transcriptome analyses of the molecular mechanisms of flower color mutation in tobacco. *BMC Genomics*. 2020;21(1):611.
89. Zheng Y, Chen Y, Liu Z, Wu H, Jiao F, Xin H, et al. Important roles of key genes and Transcription Factors in Flower Color Differences of *Nicotiana glauca*. *Genes*. 2021;12(12):1976.
90. Rausher MD. Evolutionary transitions in Floral Color. *Int J Plant Sci*. 2008;169(1):7–21.
91. Luo P, Ning G, Wang Z, Shen Y, Jin H, Li P et al. Disequilibrium of Flavonol synthase and Dihydroflavonol-4-Reductase expression Associated tightly to White vs. Red Color Flower formation in plants. *Front Plant Sci*. 2016;6.
92. Deutscher D, Meilijon I, Schuster S, Ruppin E. Can single knockouts accurately single out gene functions? *BMC Syst Biol*. 2008;2(1):50.
93. Wheeler LC, Dunbar-Wallis A, Schutz K, Smith SD. Evolutionary walks through flower colour space driven by gene expression in *Petunia* and allies (*Petuniaceae*). *Proc R Soc B Biol Sci*. 2023;290(2022):20230275.
94. Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. Contrasting frequencies and effects of *cis*- and *trans*-Regulatory mutations affecting gene expression. *Mol Biol Evol*. 2016;33(5):1131–46.
95. Benowitz KM, Coleman JM, Allan CW, Matzkin LM. Contributions of *cis*- and *trans*-Regulatory Evolution to Transcriptomic Divergence across Populations in the *Drosophila* *mojavensis* Larval Brain. *Betran E*, editor. *Genome Biol Evol*. 2020;12(8):1407–18.
96. Gates DJ, Olson BJS, Clemente TE, Smith SD. A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *lochroma*. *New Phytol*. 2018;217(3):1346–56.
97. Vimolmangkang S, Han Y, Wei G, Korban SS. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol*. 2013;13(1):176.
98. Takahashi R, Yamagishi N, Yoshikawa N. A MYB Transcription Factor Controls Flower Color in soybean. *J Hered*. 2013;104(1):149–53.
99. Streisfeld MA, Young WN, Sobel JM. Divergent selection drives genetic differentiation in an R2R3-MYB transcription factor that contributes to Incipient Speciation in *Mimulus aurantiacus*. *PLoS Genet*. 2013;9(3):e1003385.
100. Quattrocchio F, Wing JF, Leppen HTC, Koes RE. Regulatory genes Controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of Target Genes:17.
101. Farré G, Blancquaert D, Capell T, Van Der Straeten D, Christou P, Zhu C. Engineering Complex metabolic pathways in plants. *Annu Rev Plant Biol*. 2014;65(1):187–223.
102. Winkel BSJ. Metabolic channeling in plants. *Annu Rev Plant Biol*. 2004;55(1):85–107.
103. Finnegan PM, Chen W. Arsenic toxicity: the effects on plant metabolism. *Front. Physio*. 2012;3:182. <https://doi.org/10.3389/fphys.2012.00182>.
104. Zhang P, Du H, Wang J, Pu Y, Yang C, Yan R, et al. Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotechnol J*. 2020;18(6):1384–95.
105. Ruxton GD, Schaefer HM. Floral colour change as a potential signal to pollinators. *Curr Opin Plant Biol*. 2016;32:96–100.
106. Petroni K, Tonelli C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci*. 2011;181(3):219–29.
107. Willmer P, Stanley DA, Steijven K, Matthews IM, Nuttman CV. Bidirectional Flower Color and shape changes allow a second opportunity for pollination. *Curr Biol*. 2009;19(11):919–23.
108. Albert NW, Lewis DH, Zhang H, Irving LJ, Jameson PE, Davies KM. Light-induced vegetative anthocyanin pigmentation in *Petunia*. *J Exp Bot*. 2009;60(7):2191–202.
109. Nakabayashi R, Yonekura-Sakakibara K, Urano K, Suzuki M, Yamada Y, Nishizawa T, et al. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J*. 2014;77(3):367–79.
110. Butler T, Dick C, Carlson ML, Whittall JB. Transcriptome analysis of a Petal Anthocyanin Polymorphism in the Arctic Mustard, *Parrya nudicaulis*. *PLoS ONE*. 2014;9(7):e101338.
111. Sobel JM, Streisfeld MA. Flower color as a model system for studies of plant evo-devo. *Front. Plant Sci*. 2013;4:321. <https://doi.org/10.3389/fpls.2013.00321>.
112. Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, et al. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol*. 2015;207(4):1170–80.
113. Ho WW, Smith SD. Molecular evolution of anthocyanin pigmentation genes following losses of flower color. *BMC Evol Biol*. 2016;16(1):98.
114. Sapir Y, Gallagher MK, Senden E. What maintains Flower Colour variation within populations? *Trends Ecol Evol*. 2021;36(6):507–19.
115. Roguz K, Gallagher MK, Senden E, Bar-Lev Y, Lebel M, Helicz R, et al. All the colors of the Rainbow: diversification of Flower Color and Intraspecific Color Variation in the Genus *Iris*. *Front Plant Sci*. 2020;11:569811.
116. Goodrich J, Carpenter R, Coen ES. A common gene regulates pigmentation pattern in diverse plant species. *Cell*. 1992;68(5):955–64.
117. Duan HR, Wang LR, Cui GX, Zhou XH, Duan XR, Yang HS. Identification of the regulatory networks and hub genes controlling alfalfa floral pigmentation variation using RNA-sequencing analysis. *BMC Plant Biol*. 2020;20(1):110.
118. Liu C, Wang X, Shulaev V, Dixon RA. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nat Plants*. 2016;2(12):16182.
119. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 1999;16(2):266–74.
120. Ramos-Onsins SE, Puerma E, Balaña-Alcaide D, Salguero D, Aguadé M. Multi-locus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Mol Ecol*. 2008;17(5):1211–23.
121. Shoeva OYu G, AYu, Khlestkina EK. The factors affecting the evolution of the anthocyanin biosynthesis pathway genes in monocot and dicot plant species. *BMC Plant Biol*. 2017;17(2):256.
122. Wheeler LC, Walker JF, Ng J, Deanna R, Dunbar-Wallis A, Backes A et al. Transcription factors evolve faster than their structural gene targets in the Flavonoid pigment pathway. *Mol Biol Evol*. 2022;39(3).
123. Winkel-Shirley B. Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiol Plant*. 1999;107(1):142–9.
124. Wang Y, Chen S, Yu O. Metabolic engineering of flavonoids in plants and microorganisms. *Appl Microbiol Biotechnol*. 2011;91(4):949–56.
125. Choudhary N, Pucker B. Conserved amino acid residues and gene expression patterns associated with the substrate preferences of the competing

- enzymes FLS and DFR. bioRxiv. 2024;2023.11.05.565693. <https://doi.org/10.1101/2023.11.05.565693>.
126. Stafford HA. Anthocyanins and betalains: evolution of the mutually exclusive pathways. *Plant Sci.* 1994;101(2):91–8.
 127. Ballester AR, Molthoff J, de Vos R, Hekkert B, te L, Orzaez D, Fernandez-Moreno JP, et al. Biochemical and Molecular Analysis of Pink Tomatoes: deregulated expression of the gene encoding transcription factor SIMYB12 leads to Pink Tomato Fruit Color. *Plant Physiol.* 2010;152(1):71–84.
 128. Sun SS, Gugger PF, Wang QF, Chen JM. Identification of a *R2R3-MYB* gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (*Nelumbo Adans*). *PeerJ.* 2016;4:e2369.
 129. Zhao A, Cui Z, Li T, Pei H, Sheng Y, Li X, et al. mRNA and miRNA expression analysis reveal the regulation for Flower Spot Patterning in *Phalaenopsis* 'Panda'. *Int J Mol Sci.* 2019;20(17):4250.
 130. Peng J, Dong X, Xue C, Liu Z, Cao F. Exploring the Molecular Mechanism of Blue Flower Color Formation in *Hydrangea macrophylla* cv. Forever Summer. *Front Plant Sci.* 2021;12:585665.
 131. Nakatsuka A, Yamagishi M, Nakano M, Tasaki K, Kobayashi N. Light-induced expression of basic helix-loop-helix genes involved in anthocyanin biosynthesis in flowers and leaves of Asiatic hybrid lily. *Sci Hortic.* 2009;121(1):84–91.
 132. Li H, Yang Z, Zeng Q, Wang S, Luo Y, Huang Y, et al. Abnormal expression of bHLH3 disrupts a flavonoid homeostasis network, causing differences in pigment composition among mulberry fruits. *Hortic Res.* 2020;7(1):1–19.
 133. Lim SH, Kim DH, Kim JK, Lee JY, Ha SH. A Radish Basic Helix-Loop-Helix transcription factor, RsTT8 acts a positive Regulator for Anthocyanin Biosynthesis. *Front Plant Sci.* 2017;8:1917.
 134. Pál C, Papp B, Hurst LD. Highly expressed genes in yeast evolve slowly. *Genetics.* 2001;158(2):927–31.
 135. Gout JF, Kahn D, Duret L, Consortium PPG. The relationship among Gene expression, the evolution of Gene Dosage, and the rate of protein evolution. *PLOS Genet.* 2010;6(5):e1000944.
 136. Yang L, Gaut BS. Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. *Mol Biol Evol.* 2011;28(8):2359–69.
 137. Guo Y, Liu J, Zhang J, Liu S, Du J. Selective modes determine evolutionary rates, gene compactness and expression patterns in Brassica. *Plant J.* 2017;91(1):34–44.
 138. Tao W, Li R, Li T, Li Z, Li Y, Cui L. The evolutionary patterns, expression profiles, and genetic diversity of expanded genes in barley. *Front Plant Sci.* 2023;14:1168124. <https://doi.org/10.3389/fpls.2023.1168124>.
 139. Song H, Gao H, Liu J, Tian P, Nan Z. Comprehensive analysis of correlations among codon usage bias, gene expression, and substitution rate in *Arachis duranensis* and *Arachis ipaënsis* orthologs. *Sci Rep.* 2017;7(1):14853.
 140. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. Common pattern of evolution of Gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol.* 2004;21(7):1308–17.
 141. Pucker B, Walker-Hale N, Dzurlic J, Yim WC, Cushman JC, Crum A, Yang Y, Brockington SF. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented Caryophyllales, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytol.* 2024;241:471–489. <https://doi.org/10.1111/nph.19341>.
 142. Guillaume F, Otto SP. Gene Functional Trade-Offs and the evolution of Pleiotropy. *Genetics.* 2012;192(4):1389–409.
 143. Sakuta M, Tanaka A, Iwase K, Miyasaka M, Ichiki S, Hatai M, et al. Anthocyanin synthesis potential in betalain-producing Caryophyllales plants. *J Plant Res.* 2021;134(6):1335–49.
 144. Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.* 2004;40(1):22–34.
 145. Hatlestad GJ, Akhavan NA, Sunnadeniya RM, Elam L, Cargile S, Hembd A, et al. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nat Genet.* 2015;47(1):92–6.
 146. Mitsuda N, Ohme-Takagi M. Functional Analysis of Transcription Factors in *Arabidopsis*. *Plant Cell Physiol.* 2009;50(7):1232–48.
 147. Ali F, Seshasayee ASN. Dynamics of genetic variation in transcription factors and its implications for the evolution of regulatory networks in Bacteria. *Nucleic Acids Res.* 2020;48(8):4100–14.
 148. Liang M, Foster CE, Yuan YW. Lost in translation: molecular basis of reduced flower coloration in a self-pollinated monkeyflower (*Mimulus*) species. *Sci Adv.* 2022;8(37):eabo1113.
 149. Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol.* 1997;29(12):1305–12.
 150. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell.* 2018;175(7):1842–e185516.
 151. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of Cell-Type-Specific and Shared transcription factor binding sites. *Mol Cell.* 2013;52(1):25–36.
 152. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021;22(2):96–118.
 153. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol.* 2001;4(5):447–56.
 154. Pireyre M, Burow M. Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol Plant.* 2015;8(3):378–88.
 155. Lang X, Li N, Li L, Zhang S. Integrated Metabolome and Transcriptome Analysis uncovers the role of anthocyanin metabolism in *Michelia maudiae*. *Int J Genomics.* 2019;2019:e4393905.
 156. Sun X, He L, Guo Z, Xiao Z, Su J, Liu X, et al. Comparative transcriptome analyses reveal genes related to pigmentation in the petals of a flower color variation cultivar of *Rhododendron Obtusum*. *Mol Biol Rep.* 2022;49(4):2641–53.
 157. Zhang H, Tian H, Chen M, Xiong J, Cai H, Liu Y. Transcriptome analysis reveals potential genes involved in flower pigmentation in a red-flowered mutant of white clover (*Trifolium repens* L.). *Genomics.* 2018;110(3):191–200.
 158. Zhang J, Sui C, Wang Y, Liu S, Liu H, Zhang Z, et al. Transcriptome-wide analysis reveals key DEGs in Flower Color Regulation of *Hosta plantaginea* (Lam.) Aschers. *Genes.* 2020;11(1):31.
 159. Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide Sequence Database Collaboration. The sequence read Archive. *Nucleic Acids Res.* 2011;39(1):19–21.
 160. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
 161. Pucker B. Collection of color differences between plant species [Internet]. 2023. <https://github.com/bpucker/codi>.
 162. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52.
 163. Pucker B, Holtgräwe D, Rosleff Sörensen T, Stracke R, Viehöver P, Weisshaar B. A De Novo Genome Sequence Assembly of the *Arabidopsis thaliana* Accession Niederzenz-1 Displays Presence/Absence Variation and Strong Synteny. Vandepoel K, editor. *PLOS ONE.* 2016;11(10).
 164. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7.
 165. R Core Team. (2022). R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; <https://www.R-project.org/>.
 166. Wickman H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org>.
 167. Haak M, Vinke S, Keller W, Droste J, Rückert C, Kalinowski J et al. High Quality De Novo Transcriptome Assembly of *Croton tiglium*. *Front Mol Biosci.* 2018;5.
 168. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8(8):1494–512.
 169. Wheeler DL. Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 2003;31(1):28–33.
 170. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* 2005;33(2):677–80.
 171. Pucker B, Reiher F, Schilbert HM. Automatic identification of players in the Flavonoid biosynthesis with application on the Biomedical Plant *Croton tiglium*. *Plants.* 2020;9(9):1103.
 172. Rempel A, Choudhary N, Pucker B. KIPes3: automatic annotation of biosynthesis pathways. *Jul: Bioinformatics; 2022.*
 173. Pucker B. Automatic identification and annotation of MYB gene family members in plants. *BMC Genomics.* 2022;23(1):220.
 174. Pucker B, Iorizzo M. Apiaceae FNS1 originated from F3H through tandem gene duplication. *PLoS ONE.* 2023;18(1):e0280155.
 175. Price MN, Dehal PS, Arkin AP. FastTree: Computing large minimum evolution trees with profiles instead of a Distance Matrix. *Mol Biol Evol.* 2009;26(7):1641–50.
 176. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
 177. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.

178. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* 2021;19(1):232.
179. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49(W1):W293–6.
180. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–9.
181. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circIzize* implements and enhances circular visualization in R. *Bioinformatics.* 2014;30(19):2811–2.
182. Wickman H, Francois R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation [Internet]. 2022. <https://CRAN.R-project.org/package=dplyr>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.