Genetic factors explaining anthocyanin pigmentation differences

Maria Fernanda Marin Recinos[1] and Boas Pucker[1,*]

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

* corresponding author: b.pucker@tu-braunschweig.de

ABSTRACT

Background: Anthocyanins represent one of the most abundant coloration factors found in plants. Biological functions of anthocyanins range 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 across flowering plant species was performed. Literature was screened for cases in which causal genes were previously identified. Transcriptomic data sets were processed to determine the genes most likely to be responsible for color variation based on their expression pattern. The contribution of different structural and regulatory genes to the pigmentation differences was quantified. Gene expression 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 substantially more likely to explain pigmentation differences than bHLH or WD40 genes.

Conclusions: These findings support previous assumptions about the plasticity of transcriptional regulation and its importance for the evolution of novel coloration phenotypes. The particular significance of MYBs and their apparent dominant role in the specificity of the MBW complex are underlined by our findings.

INTRODUCTION

The plant kingdom is characterized by an enormous diversity of flower hues and shapes[1,2]. Some plant species maintain their brilliant colors throughout the year, while others show a constant transformation as the seasons change. The substances responsible for these colors are pigments known as 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].

Anthocyanins and other flavonoids are a group of specialized plant metabolites responsible for numerous functions beyond coloration. Additional physiological functions are 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]. Not only anthocyanins, but also flavonols are contributing to the attraction of pollinators by forming guiding signals on flowers which are invisible to the human eye[12,13]. Pigmented anthocyanins are known to be responsible for the coloration of flowers raging from red to orange to purple and to blue colors. 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[14,15] and acylation[16]. 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[6], plants exposed to high temperatures can show degradation of anthocyanins while low temperatures can increase color intensity[13].

Besides anthocyanins there are three major classes of flavonoids produced by a wide range of plant species: flavones, flavonols, and proanthocyanidins. Each of these subgroups of flavonoids have individual biological funcitions and can influence the coloration on different plant structures[17]. The characteristic cream white or pale yellow color which determined the presence of flavones and flavonols can be observed in leaves or petals of *Taraxacum officinale* “dandelion”[18], *Chrysanthemum grandiflorum* cv. Jinba[19], and *Chrysanthemum morifolium*[20]. Proanthocyanidins also called condensed tannins are colorless compounds that turn brown upon oxidation[21]. They have been studied in seeds of species such as *Arabidopsis thaliana*[22], *Brassica napus*[23], and *Ipomoea purpurea*[24].

The general pathway of the flavonoid biosynthesis (Figure 1A) is well understood and the central aglycon biosynthesis is largely conserved among land plants[3,25]. 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 naringerin, 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 oxidated through the activity of flavone synthase (FNS) and form flavones. After the hydroxylation of naringerin 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 the oxidation with flavonol synthase (FLS) or leucoanthocyanidins by the reduction with dihydroflavonol 4-reductase (DFR). In case the latter occurs, another enzyme called anthocyanidin synthase/leucoanthocyanidin dioxygenase (ANS/LDOX) catalyzes the synthesis of anthocyanidins. These anthocyanidins can be further modified through different steps including (Figure 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.

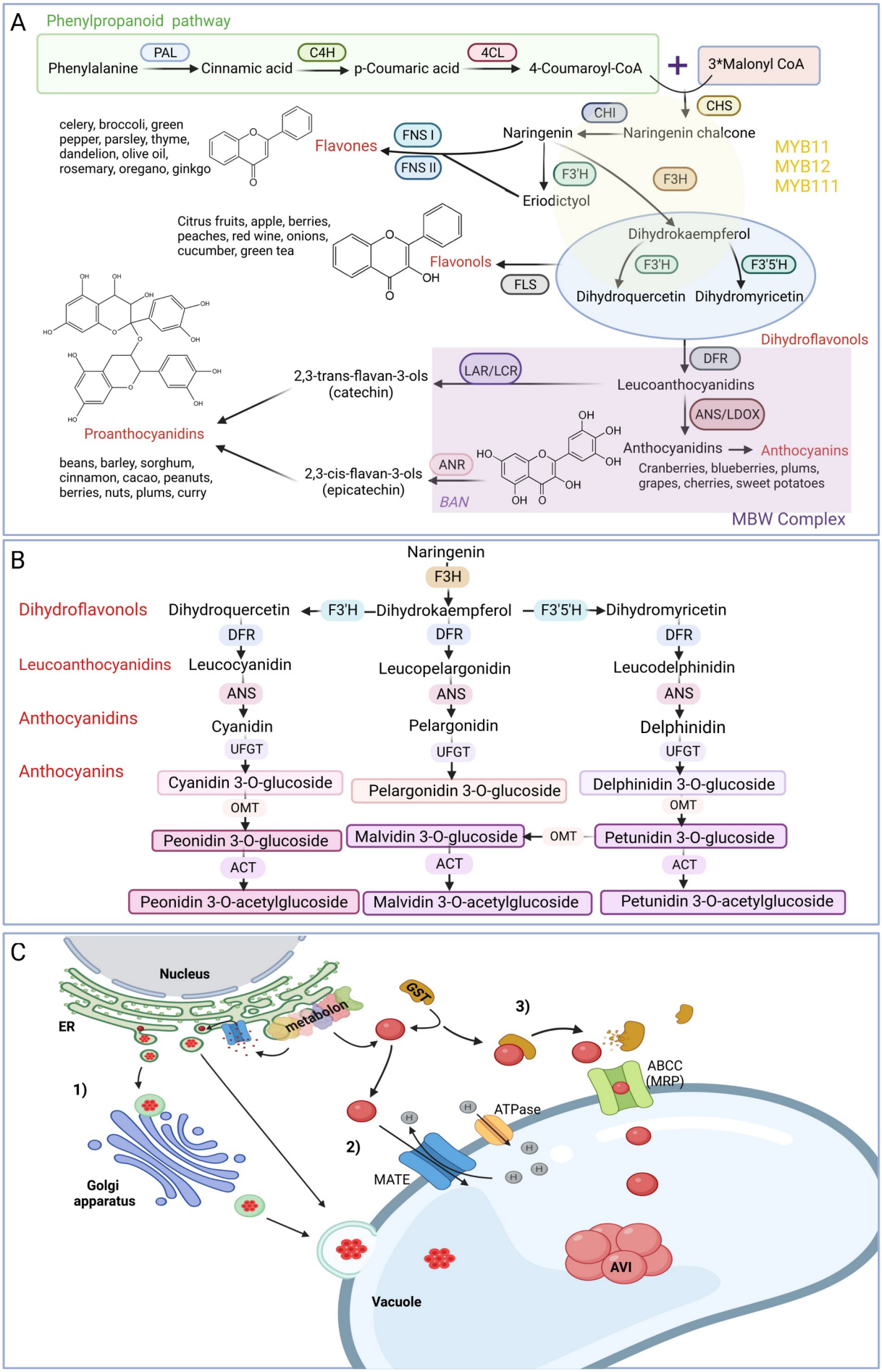
A
Phenylpropanoid pathway
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
4-Coumaroyl-CoA
3*Malonyl CoA
CHS
CHI
Naringenin chalcone
Naringenin
celery, broccoli, green pepper, parsley, thyme, dandelion, olive oil, rosemary, oregano, ginkgo
Flavones
FNS I
FNS II
F3'H
F3H
Eriodictyol
MYB11
MYB12
MYB111
Citrus fruits, apple, berries, peaches, red wine, onions, cucumber, green tea
Flavonols
FLS
Dihydrokaempferol
F3'H
F3'5'H
Dihydroquercetin
Dihydromyricetin
Dihydroflavonols
DFR
LAR/LCR
Leucoanthocyanidins
2,3-trans-flavan-3-ols (catechin)
ANS/LDOX
Proanthocyanidins
Anthocyanidins
Anthocyanins
Cranberries, blueberries, plums, grapes, cherries, sweet potatoes
ANR
BAN
beans, barley, sorghum, cinnamon, cacao, peanuts, berries, nuts, plums, curry
2,3-cis-flavan-3-ols (epicatechin)
MBW Complex
B
Naringenin
F3H
Dihydroflavonols
Dihydroquercetin
F3'H
Dihydrokaempferol
F3'5'H
Dihydromyricetin
DFR
Leucoanthocyanidins
Leucocyanidin
Leucopelargonidin
Leucodelphinidin
ANS
Anthocyanidins
Cyanidin
Pelargonidin
Delphinidin
UFGT
Anthocyanins
Cyanidin 3-O-glucoside
Pelargonidin 3-O-glucoside
Delphinidin 3-O-glucoside
OMT
Peonidin 3-O-glucoside
Malvidin 3-O-glucoside
Petunidin 3-O-glucoside
ACT
Peonidin 3-O-acetylglucoside
Malvidin 3-O-acetylglucoside
Petunidin 3-O-acetylglucoside
C
Nucleus
ER
metabolon
GST
3)
ABCC (MRP)
1)
ATPase
2)
MATE
Golgi apparatus
AVI
Vacuole

Figure 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 - anthocyanidin synthase, UFGT - UDP-glucose:flavonoid 3-O-glucosyltransferase, FLS - flavonol synthase, ANR - anthocyanidin reductase, LDOX - leucoanthocyanidin dioxygenase, FNS - flavone synthase. (C) Scheme of the different intracellular flavonoid transports mechanisms. 1. vesicle trafficking from the Endoplasmic Reticulum (ER) 2. membrane transporter throughout the Multidrug and Toxin Extrusion Transporter (MATE), and 3. Transport mediated by the Glutathione S-transferase (GST) through the ATP-binding cassette (ABCC).

Like many other specialized metabolites, flavonoids are synthesized in the cytoplasm. It is assumed that some enzymes responsible in catalyzing specific reactions in the flavonoid biosynthesis are attached to the endoplasmatic reticulum (ER) and form a metabolon[26](Figure 1C). Flavonoids produced at the ER are transported into the vacuole for storage[27–29] 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 of a range of 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[28]. Both models have in common that a mechanism 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[26,29]. The existence of these vesicles has been reported mainly in *Zea mays*[30], *A. thaliana*[31], *Vitis vinifera*[32], and *Oryza sativa*[33]. Microscopic evidence shows that these vesicular bodies are attached to the surface of the ER[32,34] 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[35]. The GST-mediated transport

model proposes that flavonoids are delivered to the tonoplast by ligandins[36,37]. These 'ligandins' are glutathione S-transferase (GST) binding and carrier proteins[38]. Evidence for the role of GST in flavonoid transport has been reported in multiple species, such as *Zea mays*[36], *A. thaliana*[37,39], *Petunia hybrida*[38], *Vitis venifera*[32], and *Prunus persica*[40]. The tonoplast-based transport mechanism involves different transmembrane channels which enable translocation of flavonoids into the vacuole[26]. 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[41]. 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[41–43].

The transcriptional activity of genes encoding enzymes of the flavonoid biosynthesis pathway is controlled by numerous transcription factors or even protein complexes comprising multiple transcription factors (Figure 1A). Genes of the flavonol and flavone biosynthesis are largely regulated by MYB11, MYB12, and MYB111[44–46]. The mechanism that regulates the expression of all structural genes in the anthocyanin biosynthesis pathway is commonly known as the MBW complex[47]. 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 combinatorical diversity[47]. After the discovery of TT8 in *A. thaliana*[48], Baudry *et al*.[49] 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[50]. 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[50].

It is assumed that the function of regulating anthocyanin biosynthesis by the MBW complex is evolutionary conserved across angiosperms[51]. Elements of the MBW complex are also involved in the control of other

biological processes in plants[52,53], which adds evolutionary constraints to all components of this complex. This diversity of functions depends mostly on the specificity of the DNA binding site of the transcription factor[54]. Additionally, it highlights the importance of TF mechanisms of action, particularly those of the MYB family, in the production of flavonoids.

Closely related plant species can differ in their anthocyanin repertoire and pigmentation pattern[55,56]. These differences can even appear between plants of the same species[1,57–59]. 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*[60], *Vitis vinifera*[61], *Hordeum vulgare*[62], *Nicotiana tabacum*[63], and *Nicotiana alata*[64].. Substrate competition between different branches of the flavonoid biosynthesis can also have an impact on the anthocyanin accumulation. For example, an increased flavonol production can lead to a pigmentation loss due to insufficient substrate for the anthocyanin biosynthesis[65]. 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[66], but previous research suggests that some genes are more often responsible for pigmentation loss than others.

This raises the question whether there are “hot spots” in the pathway that are frequently the cause for a pigmentation loss. It is generally assumed that transcriptional control is more dynamic during evolution than the biochemical function of structural genes in the flavonoid biosynthesis. Therefore, we performed a systematic comparison of anthocyanin-pigmented vs. non-pigmented varieties across flowering plant species. Literature was screened for causal genes that were previously reported as factors of pigmentation differences. Transcriptomic data sets were processed to determine the genes responsible for color variation in additional species. The contribution of different factors to the pigmentation loss was quantified and supported a crucial role of transcription factors.

2. MATERIALS AND METHODS

2.1 Extensive literature screen

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 150 full-text articles were included for an eligibility assessment during a period of seven months, from December 2021 to July 2022. Accessible articles were considered if the genetic basis of pigmentation was investigated in the respective study (Additional file 1: Table S1).

2.2 Data sources

From the 150 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*[67], *Rhododendron obtusum*[68], *Trifolium repens*[69], and *Hosta plantaginea*[70] (Additional file 3). 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 plants species were retrieved from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) (Additional file 1: Table S2) using fastq-dump[71].

2.3 Transcriptome assembly

Trimmomatic v0.39[72] 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 to enable the following assembly with Trinity[73]. Trinity v2.4.0[74] was applied for *de novo* transcriptome assembly using the previous cleaned reads as input. Trinity was run with a k-mer length of 25. Assembly statistics were computed using a previously developed python script[75] (Additional file 1: Table S3).

After completion of the assembly process, kallisto v0.44[76] was run to quantify the abundances of transcripts based on all available RNA-Seq data of the respective species. A principal components analysis (PCA) was performed using R v.4.1.3[77] with the package ggplot2 v.3.4.0[78] to inspect the variation within the data set (Additional file 2: Figures S1-S4).

2.4 Identification of Candidates Genes

In order to facilitate the identification of candidate genes, encoded peptide sequences were inferred from the transcriptome assembly using a previously established approach[79] that combines Transdecoder[80], ORFfinder[81], and ORFpredictor[82]. Knowledge-based Identification of Pathway Enzymes (KIPEs3) v0.34[83,84] was applied to identify the structural genes involved in the flavonoid biosynthesis. Flavonoid biosynthesis regulating R2R3-MYB genes were identified via MYB_annotator v.0.3[85]. A previously described BLAST-based python script[86] was deployed for the identification of additional candidates genes using bHLH, WD40, and WRKY genes associated with the flavonoid biosynthesis as baits[73]. 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[87] 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[88] using default parameters and RAxML v.8.2.12[89] with PROTGAMMA+LG+F.

2.5 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[90], circlize 0.4.15[91] and dplyr 1.1.0[92]. Genes with an adjusted p-value < 0.05 and absolute log2 fold-change > 1 were considered as differently expressed. The script used for the heatmap construction are available in our GitHub repository[73].

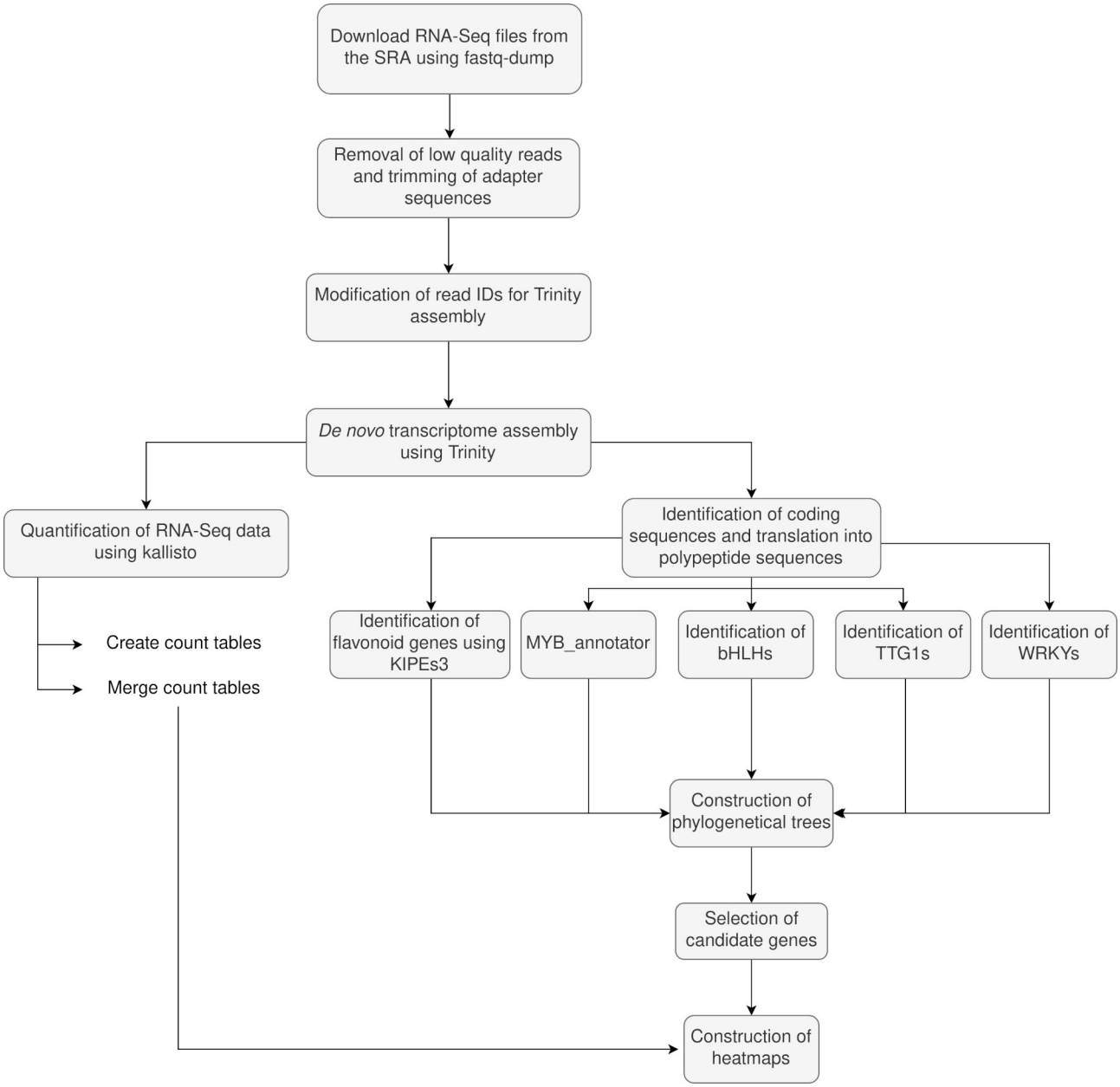

Figure 2. Flowchart representation of the methodology followed in the comparative transcriptional analysis. SRA – Sequence Read Archive, KIPEs – Knowledge-based Identification of Pathway Enzymes, MYB - Myeloblastosis, bHLH – basic helix-loop-helix, *TTG1 – Transparent Testa Glabra 1*, WRKY – WRKY DNA-binding domain.

3. RESULTS

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. In order to validate the quality of the transcriptome assemblies, a summary of the assembly statistics was generated for each species (Additional file 1: Table S3). Similarly, a PCA for every dataset was constructed based on their transcriptomic profiles to compare the similarity between samples and to identify any outliers (Additional file 2: Figures S1-S4). Phylogenetic trees for the transcription factor families MYB, bHLH, TTG1, and WRKY were constructed to assess orthologues relationships (Additional file 2: Figures S8-S17).

3.1 Michelia maudiae

A study about the flower color variation in *Michelia maudiae*[67] generated RNA-Seq data sets (Table S2) of plants of an anthocyanin-pigmented accession (red flower) and of plants of an unpigmented accession (white flower). Flower samples were taken at an early and a late developmental stage to compare not only accessions, but also developmental stages. We identified the transcript sequences of flavonoid biosynthesis genes in our transcriptome assembly of *M. maudiae*[73]. The results from our re-analyses showed high expression levels of *F3H* and *TT19* in the red flowers regardless of the development stage. The gene *ANR* showed high expression in the white flowers particularly in the late developmental stage. The gene expression of *DFR* had a slightly increased expression in the red late flowers (Figure 3A).

The genes related to the WD40 transcription factor TTG1 (Figure 3B) presented higher expression values in the white flowers independently of the developmental stage. Gene expression of *TTG1* in the red flower was considerably higher at the late developmental stage compared to the early one. Members of the transcription factor families R2R3-MYB, bHLH, WD40 and WRKY were also identified and analyzed with respect to their gene expression. Three genes related to the R2R3-MYB transcription factors associated with the flavonoid biosynthesis were detected. From those, the repressor *MYB4* showed a high expression at the late

stage of red flowers. In addition, a slight increase in expression over the different developmental stages was observed for the genes *MYB12, MYB11, MYB111,* and *MYB75, MYB90, MYB113, MYB114* (Figure 3C). Two genes of the bHLH transcription factor family associated with the regulation of the flavonoid biosynthesis were discovered. One gene is related to *LhbHLH1* and the other one to *TT8* (Figure 3D). The expression of the genes related to *TT8* were not particularly different between the flower stages. On the contrary, the genes related to *LhbHLH1* showed a higher expression in samples taken from the red flowering accession at the early stage. Furthermore, in the transcription family WRKY (Figure 3E) four different genes were identified as putative orthologs of the anthocyanin regulators reported in *Malus domestica and Pyrus spp.* Orthologs of the genes MdWRKY40, and PyWRKY31 presented the higher expression in samples of the red flowers compared to the white flowers at the early developmental stage. In summary, there are several differences in gene expression like the strong expression of *TT19* in the red samples and the almost complete absence of expression of this gene in the white samples which suggest that this step could be the bottleneck in the anthocyanin accumulation which ultimately results in white flowers. However, the gene expression differences of other genes in the anthocyanin biosynthesis suggest that the causal gene is more likely to be a transcription factor that activates multiple structural genes.

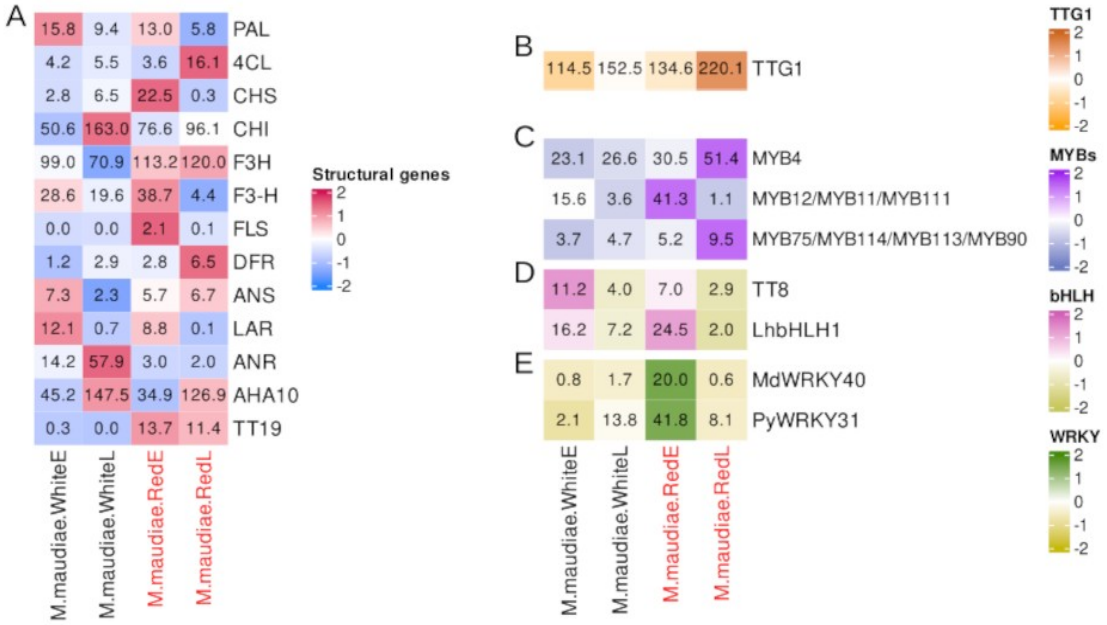

Figure 3. Heatmaps presenting gene expression values of white and red *Michelia maudiae* flowers and their respective developmental stages, early (E) and late (L). (A) Structural genes, (B) *TTG1*, (C) MYBs, (D) bHLH, and (E) WRKY. An isoform resolved version of this heatmap is available in Additional file 2: Figure S5.

3.2 *Rhododendrom obtusum*

A study about the flower coloration of *Rhododendron obtusum*[68] generated a series of RNA-Seq datasets that cover five developmental stages of two varieties. The variety "Yanzhi Mi" characterized by a bright pink color and the wild type variety "Dayuanyangjin" which presents white petals with pink stripes. The authors of this study state that derivatives of cyanidin, peonidin, and pelargonidin might be responsible for the pink color of mutant petals and that the S2 stage was the key stage of flower color formation. Since no reference sequence is publicly available, a *de novo* transcriptome assembly of *R. obtusum* was generated for the identification of transcripts belonging to flavonoid biosynthesis genes[73].

Results of the re-analyses showed high expression values in stage 2 of the white variety in genes *AHA10, UGT72L1, ANR,* and *A3GT* in comparison to the pink flowers at that same stage. On the contrary, the pink flowers showed high expression for the genes *4CL, LAR,* and *MATE* at stage 2. No other particular difference between the two varieties was observed in the other stages with the exception of *F3'5'H* (Figure 4A).

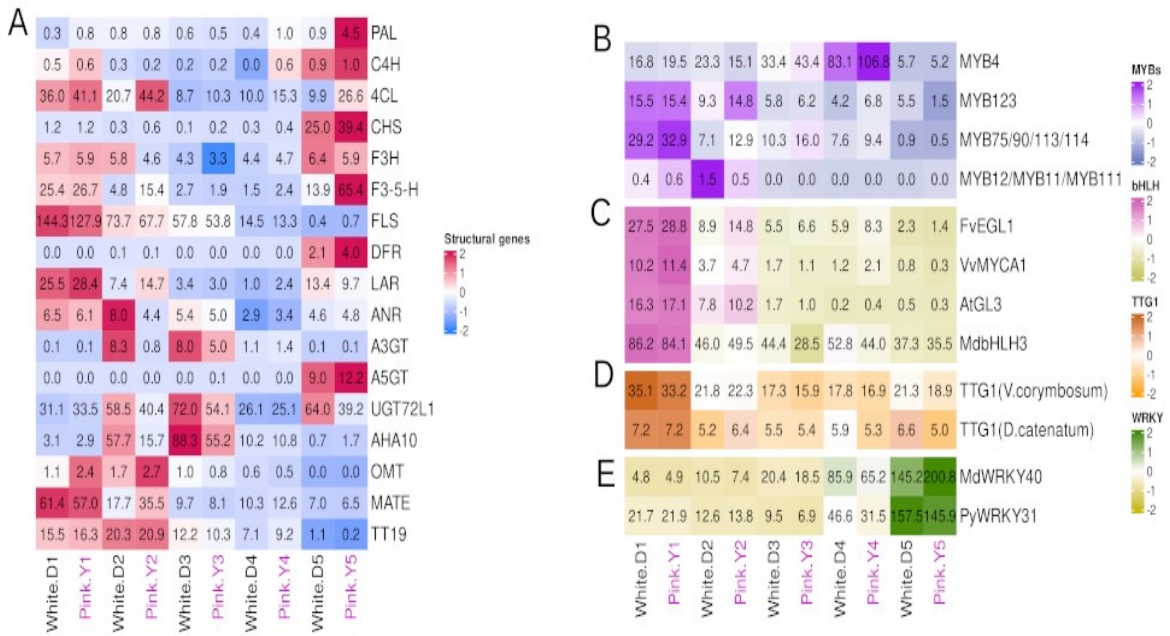

Figure 4. Heatmaps presenting gene expression values of white flowering *Rhododendron obtusum* var. *'Dayuanyangjin'* in different developmental stages: closed buds (D1), buds showing color at the top (D2), the initial flowering stage (D3), the full flowering stage (D4) and the last flowering stage (D5) and pink flowering *Rhododendrom obtusum* var. *'Yanzhi Mi'* with the different flower stages: closed buds (Y1), buds showing color at the top (Y2), the initial flowering stage (Y3 stage), the full flowering stage (Y4 stage) and the last flowering stage (Y5) (A) Structural genes, (B) *TTG1*, (C) MYBs, (D) bHLH, and (E) WRKY. An isoform resolved version of this heatmap is available in Additional file 2: Figure S6.

Two *TTG1* genes belonging to two different clades of the species *Vaccinium corymbosum* and *Dendrobium catenatum* were identified. These genes did not show a clear difference in gene expression between the pink and white samples at any developmental stage (Figure 4B). Three genes related to the *A. thaliana* transcription factors *AtMYB75, AtMYB123,* and *AtMYB4* (Figure 4C) were identified. According to Sun *et al*., S2 is the developmental stage where the difference in pigmentation becomes visible. This analysis revealed a difference in expression of the structural genes in S2 samples. It is also interesting to notice a slight difference in the expression of genes related to the transcription factor families MYB and bHLH between the same two S2 groups. This observation suggests a possible involvement of MYB and bHLH in the regulation divergence (Figure 4D). The WRKY transcription factor family shows no particular difference in expression between the varieties at any developmental stage (Figure 4E). In summary, it seems that a slightly higher activity of *MYB75* and *FvEGL1/AtGL3* orthologs could be responsible for the strong pigmentation of the pink cultivar.

3.3 *Trifolium repens*

Trifolium repens commonly know as "white clover" is a perennial herbaceous plant native to Europe and western Asia. This plant is characterized by its white flower, however plants with red or pinkish flowers can be seen in rare cases. Zhang *et al.*, 2018 observed that those red flowering clover varieties produce white flowers under shaded conditions. Therefore, they performed an RNA-Seq analysis on the anthocyanin biosynthesis of a red clover variety to identify the genes associated with the flower color. In their results eight candidates genes were identified. Most of them were down-regulated in plants under shaded conditions (*CHS, F3'H, F3'5'H, ANS, UFGT, and DFR*) and only two genes were found to be up-regulated (*LAR* and *FLS*)[69].

A *de novo* transcriptome assembly of *T. repens* was generated for the identification of transcripts belonging to flavonoid biosynthesis genes[73]. The results of our transcriptome re-analysis (Figure 5A) confirmed the low expression levels of the structural genes *CHS, F3'H, F3'5'H, DFR, ANS, TT19, and UFGT,* for the white flowers in comparison with the red flowers. However, these white flowers presented relatively high

expression of the genes *4CL*, *AHA10*, and *TT15*. Additionally, we analyzed the expression levels of members of the MYB, bHLH, *TTG1,* and WRKY transcription factor families.

One ortholog of the *A. thaliana MYB75* was identified (Figure 5B). No distinct pattern between the varieties was observed suggesting MYBs are not likely to be responsible for the color difference. By contrast, the family bHLH (Figure 5C) presented differences between the red and white samples, particularly the unigenes related to the transcription factors *Lonicera japonica GL3* and *Arabidopsis EGL3*. This implies that a possible inactivation of bHLH due to lack of light exposure could be associated with the loss of pigmentation in the white flowers of *T. repens*. Concerning the transcription factor *TTG1* (Figure 5D) no substantial difference in expression was identified between the two groups. In theWRKY transcription factor family (Figure 5E) the red flowers presented higher expression levels, specially the genes related to *M. domestica WRKY40* and *Pyrus spp. WRKY31.* In summary, higher bHLH expression in light exposed clover plants, especially of the encoding genes of TFs *EGL3* and *GL3,* appears to be the most likely explanation for the generally increased expression levels of various structural genes.

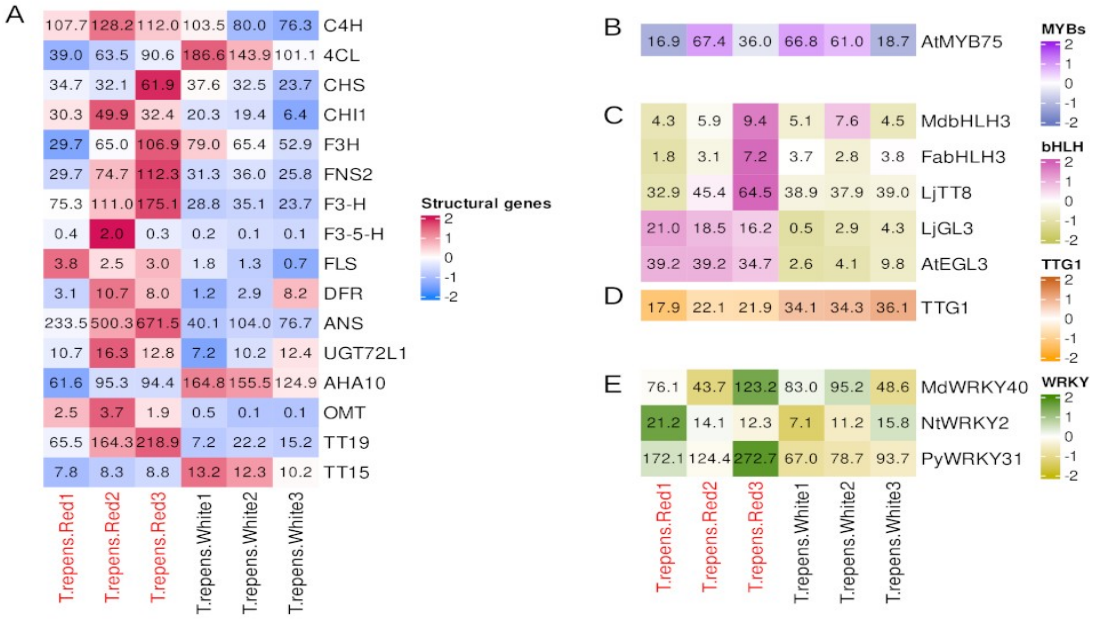

Figure 5. Heatmaps presenting gene expression values of red flowering *Trifolium repens* and white flowering *T. repens exposed to shaded conditions*. (A) Structural genes, (B) MYBs (C) bHLHs, (D) *TTG1*, and (E) WRKYs. An isoform resolved version of this heatmap is available in Additional file 2: Figure S7.

3.4 Genetic hot spots responsible for anthocyanin pigmentation differences

Based on a total of 150 analyzed studies (Additional Table S1) including reanalyzed RNA-Seq experiments, we determined the genes most likely to be responsible for color variation between accessions in each of these species (Figure 6). Pigmentation was considered to be the wild type state, while absence of pigmentation was assumed to be the result of a mutation. We identified 32 research articles that report upregulated structural genes in the anthocyanin biosynthesis as the cause of color difference between pigmented and unpigmented accessions. Additionally, 11 events of non-activated or down-regulated structural anthocyanin biosynthesis genes were reported in the literature. Moreover, 86 different studies propose a transcription factor might be responsible for differences in pigmentation. Many of these reports named a specific transcription factor. In total, 41 MYBs, six bHLHs, two TTG1 homologs, one bZIP, one WRKY were presented as the causal gene for pigmentation differences. The remaining 33 cases are likely due to the action of multiple transcription factors or caused by TFs that activate the components of the MBW complex. Five reports presented up-regulated genes that encode proposed intracellular transporters of anthocyanins as best candidates. It was not possible to determine the causal gene in 16 of the analyzed studies.

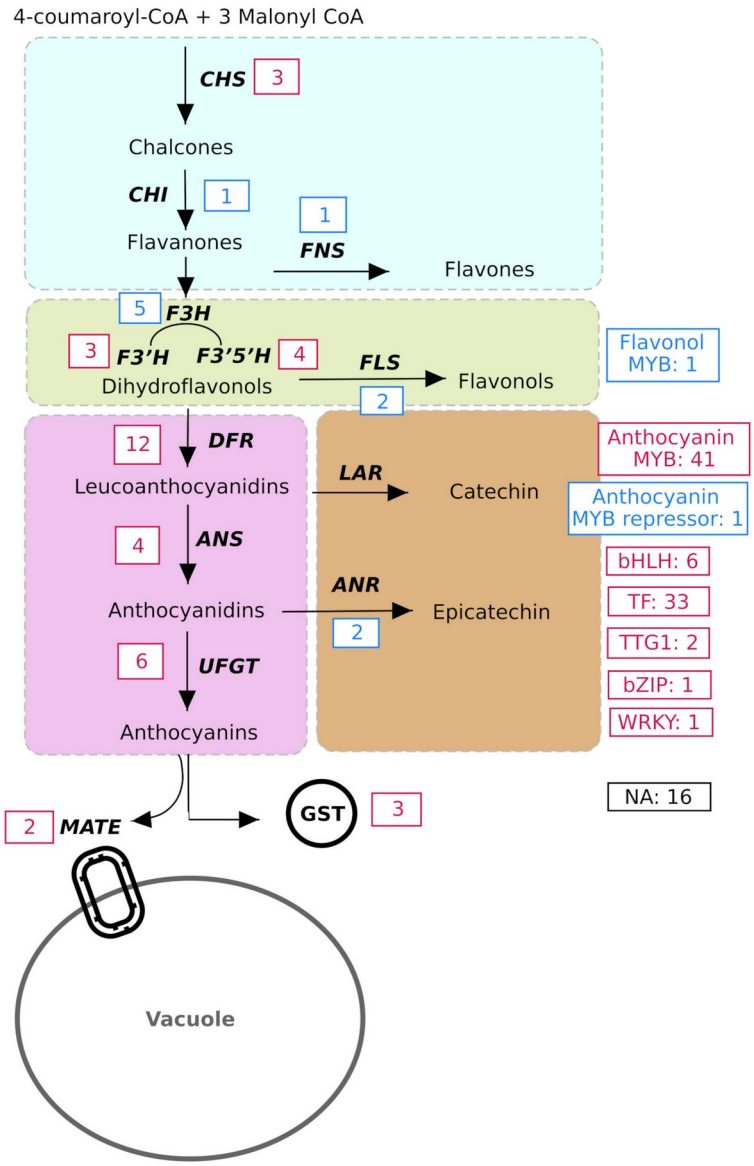

Figure 6. Schematic representation of the flavonoid biosynthesis pathway with emphasis on the number of genes responsible for color difference according to a systematic literature analysis and re-analyses of RNA-Seq data sets. The anthocyanin-pigmented accession is set as reference when determining up- and down-regulation. Red boxes and blue boxes indicate the number of upregulated and downregulated genes, respectively. Flavonoids were divided into four groups that are indicted 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, GST – Glutathione S-tranferase, MATE – Multidrug And Toxin Extrusion.

4. 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 e.g. to indicate which flowers have already been visited by insects[93] Flavonoids are also known to protect against UV radiation[94], drought[95], and cold stress [96,97]. Previous studies have identified numerous genes involved in anthocyanin biosynthesis in different plant species (Additional file 1: Table S1). We present an aggregated analysis of the most likely candidates genes responsible for color variation. This analysis harnessed numerous public RNA-Seq data sets that enable a direct comparison of differently pigmented accessions.

4.1 Michelia maudiae

M. maudiae is a perennial tree belonging to the family Magnoliaceae. Lang *et al.*[67] conducted an RNA-Seq experiment to investigate the flower color variation in *M. maudiae.* Early and late developmental stages of the white and red flowers were sampled. The original analysis did not reveal an individual gene responsible for the pigmentation difference. Our reanalysis revealed high expression of the known anthocyanin transporter gene *TT19 (GST)*[39] in the red flowers while there was no substantial gene expression in the white flowers. This suggests that a variation in the promoter of the *TT19* ortholog or its regulators could prevent the anthocyanin accumulation in the white flowering cultivar. Variations disrupting this gene were previously reported as blocks in the anthocyanin accumulation[37,38,98]. An established hypothesis states that this protein protects anthocyanins during transport to the central vacuole[98]. In addition, the structural genes *CHI, CHS, F3H, DFR and ANR* presented the highest expression levels in the red flowers compared to their expression in the white ones. It is unlikely that five structural genes would be up-regulated independently and more likely that a transcription factor is the causal factor in this scenario. Furthermore, we identified two transcription factor genes, one ortholog related to the repressor *AtMYB4*[99], and one ortholog of *LhbHLH1* known to be involved in the anthocyanin biosynthesis in *Lilium spp*[100]. Clear differences in the expression of unigenes related to the bHLH transcription factor were determined. In red flowers, genes related to *LhbHLH1* showed higher expression levels compared to white flowers.

Additionally, the higher expression of WRKY transcription factor-related genes in red 'Rubellis' flowers compared to white flowers is of particular interest. Similar results were reported by Liu *et al*. in 2020[101]. The authors reported up-regulation of bHLH and WRKY transcription factor-related genes only at the early stage of flowers belonging to the species *Michelia crassipes*. According to their claims, the reason why bHLH and WRKY genes appear to be up-regulated in the early stages of flower development could be due to their importance in petal color transition during flowering. Recent studies hypothesize that the regulation of the production of anthocyanin pigments is not entirely handled by the MBW complex, but that transcription factors such as those belonging to the WRKY family may also play a role in the regulation and transport of these pigments[102–104].

4.2 Rhododendrom obtusum

R. obtusum commonly known as Japanese azaleas is a small shrub, characterized by its pink flowers. Sun *et al*.[68] provided a starting point into the genetic mechanisms underlying the flower color divergence in the *R. obtusum* species by generating an RNA-Seq data set. They analyzed the gene expression in two differently pigmented varieties of *R. obtusum* in five flower developmental stages. The variety “Yanzhi Mi” characterized for a bright pink color and the wild type variety “Dayuanyangjin” which presents white petals with pink stripes. Their results indicated a high gene expression in *CHS, CHI, F3Hs,* and *F3'H* in the pink variety in stage 2 of flower development[68]. According to the authors, developmental stage 2 was considered as the key point of divergence in flower pigmentation between 'Yanzhi Mi' variety (pink flowers) and the 'Dayuanyangjin' variety (white flowers). Our results showed a slight difference in gene expression mainly of *FLS* at stage 2 of white flowers, which could indicate a channeling of substrate towards flavonol production. High activity of the flavonol biosynthesis is consistent with the absence of red anthocyanins in the flowers as previously reported in numerous species[65,105,106]. Moreover, an increased expression of the transporter gene *MATE* (*TT12*) in the pink flowers could explain the anthocyanin accumulation starting at the stage 2 of flower development[43]. With respect to the variety 'Yanzhi Mi' an increase in the expression of unigenes related to *MYB75* and *GL3* of the bHLH family was identified. Our findings support the results of a recent study on *Rhododendron pulchrum* species[107], which presented a comparative transcriptome analysis between three varieties with different phenotypes in flower color. Their results also determined a high expression of a gene encoding the FLS enzyme, increasing the flavonol content at the expense of

anthocyanin formation in the variety with white flowers. At the same time, this variety showed increased expression of *MYB5*, which they claimed negatively regulates anthocyanin synthesis[107].

4.3 Trifolium repens

Trifolium is a worldwide spread genus belonging to the Fabaceae family. Adaptation to different environments has made this herbaceous plant to present a great variety of phenotypes, especially variations in the color pattern of its leaves and flowers[108]. The study by Zhang *et al.*[69] concluded that low expression of *CHS* and *F3'H* in the shaded plants likely limits the formation of dihydrokaempferol, which results in a blockage in the production of the red-magenta cyanidin pigment.

Similarly, our results revealed an increase in expression of the F3’H encoding gene in the red flowers. This upregulation of *F3’H* might suggest a redirection of substrate towards dihydroquercetin, which promotes the production of anthocyanins, specifically of the cyanidin class that results in the red coloration. High expression of *ANS* is supporting this hypothesis. Despite being a critical component in the substrate contribution of the flavonoid biosynthetic pathway, the F3'H enzyme is frequently overlooked in discussions of DFR and FLS specificity. For instance, the expression of *F3'H* in grapevines increases when grapes turn red during veraison[61]. Similar results were observed in *Fragaria vesca* and *Vaccinium corymbosum* with an increase of *F3’H* transcript levels during fruit ripening [109,110]. Additionally, a clear difference in bHLH expression was observed, especially in *EGL3* and *GL3*, between flowers exposed to light (red flowers) and those kept in shade (white flowers). Flowers grown under shaded conditions showed a low expression of these bHLHs. Similar results were presented in a study by Hong *et al*.[111]. They performed a transcriptomic analysis in *Chrysanthemum* to identify the effects of light intensity on anthocyanin synthesis. The results presented by Hong *et al*. showed changes in the coloration of ray florets when their capitula were exposed to different shading conditions. Those capitula that were exposed to normal light developed into pink flowers while those that were exposed to dark conditions developed into white or pale pigmented flowers, demonstrating the effect that light has on the accumulation of anthocyanins in plant tissues. However, the results of transcriptomic analysis also identified that the genes encoding the bHLH protein showed low gene expression in those plants incubated in the shade.

Research involving different light conditions and anthocyanin synthesis has been a subject of discussion for many years. Early studies proposed that the transcription factor bZIP, together with the MBW complex, would regulate the anthocyanin formation in different plant tissues in response to light conditions [112,113]. Now, it is known that the molecular mechanism behind the impact of shade on plant color production is more complex and depends on the type of plant and the duration of the light exposure[114]. Other important factors are phytohormones and HY5, a TF of the bZIP family, which positively regulates the expression of genes related to anthocyanin synthesis upon light exposure [115,116]. The evident inactivity of bHLH in the flowers of plants growing in the shade might be responsible for the lack of *F3'H* activity that lead to the absence of pigmentation in the flowers of *T. repens.* Similar results have been observed in purple broccoli (*Brassica oleraceae*) when cultivated in the shade[117].

4.4 DFR more frequently affected than ANS

DFR activates the conversion of dihydroflavonols to leucoanthocyanidins, which is often considered as the first committed step of the anthocyanin biosynthesis. We observed that *DFR* is more likely to harbour a disruptive mutation 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[118]. 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 [119].

A study published by Rausher *et al*. [120] explains how the evolutionary rate of enzymes depends on their location in a pathway. Upstream enzymes evolve at a slower rate, because they control the flux to latter steps in the pathway[60,121]. This theory states that enzymes positioned at the beginning of a pathway are more likely to have a greater impact on the overall fitness in a species and should display slower changes[122–124]. According to the Arabidopsis Information Resource (TAIR), the length of the coding sequence (CDS) of *DFR* 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*. According to theories of metabolic regulation, it is evolutionary beneficial to have changes of such magnitude in genes acting in the upper part of a biosynthesis pathway in order to avoid a waste of energy and resources by pushing substrate into a blocked pathway[125,126].

4.5 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[127]. Metabolically, this can occur when two enzymes or transporters compete for the same or very similar substrates [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. 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, 11 studies determined that the low expression of *DFR* is responsible for the color contrast between unpigmented tissues and those that show anthocyanin pigmentation. Only two studies identified 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[128]. A recent study identified a flavonol biosynthesis regulating MYB as the most frequently affected gene in pigmentation pattern change[123]. 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 most 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 resonsible genes. Once the anthocyanin biosynthesis is broken, selection against additional mutations in the anthocyanin biosynthesis genes might be weak or completely absent. This could result in more 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 based on a more comprehensive gene sets could benefit the accurate identification of causal genes in color difference studies.

4.6 Transcription factor variations appears 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. The anthocyanin biosynthesis is even considered a model system for transcriptional control in eukaryotes. Previous studies identified transcriptional activation of different R2R3-MYBs[129–132] and bHLHs[100,133,134] as cause for the increase in anthocyanin levels. Our results showed that transcription factors were three times more often reported as causal factor of color differences than structural genes. While we cannot rule out the possibility that structural genes can also accumulate mutations 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[126]. Wheeler *et al*.[123] showed that transcription factors presented lower levels of gene expression compared to their targeted structural genes. This might suggest a negative correlation between evolution rate and gene expression.

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 thus 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[135], which do not accumulate anthocyanins[128,136].

The frequency of mutations in TF 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 likely to mutate than non-TF genes[137]. Another study examined the frequency of point mutations in

TFs genes in *Eschericchia coli* and found that these genes were more vulnerable to harmful mutations that resulted in significant changes in gene expression than non-TF genes[138]. Given that their activity covers a wide range of functions, transcription factors can regulate and affect the expression of structural genes without the need of high gene expression levels. 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.* The results of their experiments concluded that despite the similar gene expression of *PELAN* in both strong pigmented and low pigmented cultivars, the difference in phenotype was due to a mistranslation of the *PELAN* mRNA [139].

The number of TFs required to regulate a specific enzyme-encoding gene is a complex and dynamic process that is influenced by various factors[140], including the complexity of the regulatory region[141], the stage of development or cell type[142], and environmental factors[143]. For example only about ten *R2R3-MYBs* are known to bind specific DNA sites related to the regulation of flavonoid biosynthesis pathway in *A. thaliana*[46,144]. In a review study, Feller *et al.*[145] performed a comparative analysis of the transcription factor families MYB and bHLH. In their study, 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. Multiple bHLH proteins contain a similar ligand-binding domain targeting different enzyme encoding genes[145]. A greater proportion of MYBs were recognized to be responsible for the regulation of flavonoid biosynthesis genes in comparison to bHLHs[146]. Our results align with this observation, because MYBs were presented as the causal gene of color differences in 41 cases, while bHLHs were only reported as crucial factor in six cases. If bHLHs are more often involved in other processes, their loss would be more detrimental which makes it less likely.

5. CONCLUSION

A systematic literature screen supported the assumption that transcription factors are the most frequently observed blocks in the anthocyanin accumulation. The MYB and bHLH components of the MBW complex are dominating as blocks in the anthocyanin accumulation when comparing differently pigmented

accessions. The degree of transcription factor specialization for a certain pathway seems to determine the frequency of their implication in color differences. According to our results, MYBs are most often responsible for the difference in anthocyanin content, followed by bHLH, WRKY, and TTG1. 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.

Additional files

Additional file 1:

Table S1. List of assessed species in systematic literature screen.

Table S2. List of analyzed species with their respective accession numbers.

Table S3. Statistics about the quality of the different Trinity assemblies.

Table S4. List of candidate genes for *Michelia maudiae*.

Table S5. List of candidate genes for *Rhododendron obtusum*.

Table S6. List of candidate genes for *Trifolium repens*.

Table S7. List of candidate genes for *Hosta plantaginea*.

Additional file 2:

Figure S1. Principal Component Analysis (PCA) *Michelia maudiae*.

Figure S2. Principal Component Analysis (PCA) *Rhododendron obtusum*.

Figure S3. Principal Component Analysis (PCA) *Trifolium repens*.

Figure S4. Principal Component Analysis (PCA) *Hosta plantaginea*.

Figure S5. Complete isoform heatmaps *Michelia maudiae*.

Figure S6. Complete isoform heatmaps *Rhododendron obtusum*.

Figure S7. Complete isoform heatmaps *Trifolium repens*.

Figure S8. Phylogenetic trees representing the relationships of MYB transcription factors involved in the flavonoid biosynthesis pathway between *Michelia maudiae* and corresponding orthologs in *Arabidopsis thaliana.*

Figure S9. Phylogenetic trees representing the relationships of transcription factors involved in the flavonoid biosynthesis pathway between *Michelia maudiae* and corresponding orthologs in different plant species.

Figure S10. Phylogenetic trees representing the relationships of *MYB4* and *MYB75* transcription factors involved in the flavonoid biosynthesis pathway between *Rhododendron obtusum* and corresponding orthologs in *Arabidopsis thaliana.*

Figure S11. Phylogenetic trees representing the relationships of *MYB12* and *MYB123* transcription factors involved in the flavonoid biosynthesis pathway between *Rhododendron obtusum* and corresponding orthologs in *Arabidopsis thaliana.*

Figure S12. Phylogenetic trees representing the relationships of transcription factors involved in the flavonoid biosynthesis pathway between *Rhododendron obtusum* and corresponding orthologs in other plant species.

Figure S13. Phylogenetic trees representing the candidate WRKYs in *Rhododendron obtusum* related to *Malus domestica* and *Pyrus spp.*

Figure S14. Phylogenetic trees representing the relationships of MYB and bHLH transcription factors involved in the flavonoid biosynthesis pathway between *Trifolium repens* and corresponding orthologs in different plant species.

Figure S15. Phylogenetic trees representing the relationships of TTG1 and WRKY transcription factors involved in the flavonoid biosynthesis pathway between *Trifolium repens* and corresponding orthologs in different plant species.

Figure S16. Phylogenetic trees representing the relationships of MYB and bHLH transcription factors involved in the flavonoid biosynthesis pathway between *Hosta plantaginea* and corresponding orthologs in different species.

Figure S17. Phylogenetic trees representing the relationships of TTG1 and WRKY transcription factors involved in the flavonoid biosynthesis pathway between *Hosta plantaginea* and corresponding orthologs in different species.

Additional file 3

Investigation of molecular mechanisms underlying pigmentiation differences in *Hosta plantaginea*

References

1. Alkema J, Seager SL. The chemical pigments of plants. J Chem Educ. 1982 Mar;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 Aug 1;39(8):619–38.

3. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 2008 May;54(4):733–49.

4. Grotewold E. The Genetics and Biochemistry of Floral Pigments. Annu Rev Plant Biol. 2006 Jun 1;57(1):761–80.

5. Grotewold E. The science of flavonoids. New York: Springer; 2006. 273 p.

6. Burns KC, Dalen JL. Foliage color contrasts and adaptive fruit color variation in a bird-dispersed plant community. Oikos. 2002 Mar;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. Chutipaijit 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 Jan 1;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 Nov 1;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 Oct 1;26(2):771–4.

12. Samanta A, Das G, Das SK. Roles of Flavonoids in Plants. 2011;6(1):12–35.

13. 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 Nov;134(6):1335–49.

14. 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 May 18;282(20):32–41.

15. Tohge T, Yonekura-Sakakibara K, Niida R, Watanabe-Takahashi A, Saito K. Phytochemical genomics in *Arabidopsis thaliana*: A case study for functional identification of flavonoid biosynthesis genes. Pure Appl Chem. 2007 Jan 1;79(4):811–23.

16. 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.

17. Thompson WR, Meinwald J, Aneshansley D, Eisner T. Flavonols: Pigments Responsible for Ultraviolet Absorption in Nectar Guide of Flower | Science. 1972;177.

18. 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.

19. Chen SM. The identification of flavonoids and the expression of genes of anthocyanin biosynthesis in the chrysanthemum flowers. Biologia plantarum. 2012;56(3):458–64.

20. 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 Sep;166:1109–20.

21. Constabel CP. Molecular Controls of Proanthocyanidin Synthesis and Structure: Prospects for Genetic Engineering in Crop Plants. J Agric Food Chem. 2018 Sep 26;66(38):9882–8.

22. Dixon RA, Xie D, Sharma SB. Proanthocyanidins – a final frontier in flavonoid research? New Phytol. 2005 Jan;165(1):9–28.

23. Leung J, Fenton TW, Mueller MM, Clandinin OR. Condensed Tannins Of Rapeseed Meal. J Food Sci. 1979 Sep;44(5):1313–7.

24. 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 Jan 19;49(4):641–54.

25. Winkel-Shirley B. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiol. 2001 Jun 1;126(2):485–93.

26. Grotewold E, Davies K. Trafficking and Sequestration of Anthocyanins. Nat Prod Commun. 2008 Aug;3(8).

27. 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 Jan 1;90(1):7–18.

28. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. Trends Plant Sci. 2015 Sep;20(9):576–85.

29. Pucker B, Selmar D. Biochemistry and Molecular Basis of Intracellular Flavonoid Transport in Plants. Plants. 2022 Jan;11(7):963.

30. 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 May;10(5):721–40.

31. 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 Jan;3(1):78–90.

32. 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.

33. 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 Apr;10(4):685.

34. 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 Dec 1;145(4):1323–35.

35. Zhao J, Dixon RA. The ‘ins’ and ‘outs’ of flavonoid transport. Trends Plant Sci. 2010 Feb;15(2):72–80.

36. Marrs KA, Alfenito MR, Lloyd AM, Walbot V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. Nature. 1995 Jun;375(6530):397–400.

37. 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 Aug 1;123(4):1561–70.

38. 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 Jul 1;10(7):1135–49.

39. 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.

40. 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 Sep;107(5):1320–31.

41. 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.

42. 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 Apr 1;13(4):853–71.

43. 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 Jun 1;19(6):2023–38.

44. Mehrtens F, Kranz H, Bednarek P, Weisshaar B. The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiol. 2005 Jun;138(2):1083–96.

45. Stracke R, Ishihara H, Huep 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.

46. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010 Oct 1;15(10):573–81.

47. Ramsay NA, Glover BJ. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 2005 Feb;10(2):63–70.

48. 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 Oct 1;12(10):1863–78.

49. 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.

50. 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 Mar;53(5):814–27.

51. Zhang B, Hülskamp M. Evolutionary Analysis of MBW Function by Phenotypic Rescue in *Arabidopsis thaliana*. Front Plant Sci. 2019;10.

52. Symonds VV, Hatlestad G, Lloyd AM. Natural Allelic Variation Defines a Role for ATMYC1: Trichome Cell Fate Determination. PLOS Genet. 2011 Jun 9;7(6).

53. Li Y, Liu X, Cai X, Shan X, Gao R, Yang S, et al. Dihydroflavonol 4-Reductase Genes from Freesia hybrida Play Important and Partially Overlapping Roles in the Biosynthesis of Flavonoids. Front Plant Sci. 2017;8.

54. Lai Y, Li H, Yamagishi M. A review of target gene specificity of flavonoid R2R3-MYB transcription factors and a discussion of factors contributing to the target gene selectivity. Front Biol. 2013 Dec 1;8(6):577–98.

55. Streisfeld MA, Rausher MD. Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. New Phytol. 2009 Aug;183(3):751–63.

56. Smith SD, Rausher MD. Gene Loss and Parallel Evolution Contribute to Species Difference in Flower Color. Mol Biol Evol. 2011 Oct 1;28(10):2799–810.

57. Almeida J, Carpenter R, Robbins TP, Martin C, Coen ES. Genetic interactions underlying flower color patterns in *Antirrhinum majus*. Genes Dev. 1989 Jan 11;3(11):1758–67.

58. Onozaki T, Mato M, Shibata M, Ikeda H. Differences in flower color and pigment composition among white carnation (*Dianthus caryophyllu* L.) cultivars. Sci Hortic. 1999 Dec;82(1–2):103–11.

59. Zufall RA. The Genetic Basis of a Flower Color Polymorphism in the Common Morning Glory (*Ipomoea purpurea*). J Hered. 2003 Nov 1;94(6):442–8.

60. 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.

61. 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 Dec;7(1):12.

62. 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 Oct 5;11(10):e0163782.

63. 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 Dec;21(1):611.

64. 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 alata*. Genes. 2021 Dec 10;12(12):1976.

65. 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 Jan 13;6.

66. Deutscher D, Meilijson I, Schuster S, Ruppin E. Can single knockouts accurately single out gene functions? BMC Syst Biol. 2008 Jun 18;2(1):50.

67. 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 Nov 3;2019:1–14.

68. 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 Apr;49(4):2641–53.

69. 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 May;110(3):191–200.

70. 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 Jan;11(1):31.

71. Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide Sequence Database Collaboration. The Sequence Read Archive. Nucleic Acids Res. 2011 Jan 1;39(1):19–21.

72. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

73. Pucker B. Collection of color differences between plant species. 2023. Available from: https://github.com/bpucker/codi

74. 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 Jul;29(7):644–52.

75. 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. Vandepoele K, editor. PLOS ONE. 2016 Oct 6;11(10).

76. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 May;34(5):525–7.

77. R Core Team (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; Available from: https://www.R-project.org/.

78. Wickman H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org

79. 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.

80. 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 Aug;8(8):1494–512.

81. Wheeler DL. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 2003 Jan 1;31(1):28–33.

82. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 2005 Jul 1;33(2):677–80.

83. Pucker B, Reiher F, Schilbert HM. Automatic Identification of Players in the Flavonoid Biosynthesis with Application on the Biomedicinal Plant Croton tiglium. Plants. 2020 Sep;9(9):1103.

84. Rempel A, Choudhary N, Pucker B. KIPEs3: Automatic annotation of biosynthesis pathways. Bioinformatics; 2022 Jul.

85. Pucker B. Automatic identification and annotation of MYB gene family members in plants. BMC Genomics. 2022 Mar 19;23(1):220.

86. Pucker B, Iorizzo M. Apiaceae FNS I originated from F3H through tandem gene duplication. PLOS ONE. 2023 Jan 19;18(1).

87. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol. 2009 Jul 1;26(7):1641–50.

88. 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 Jan 1;32(1):268–74.

89. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–3.

90. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016 Sep 15;32(18):2847–9.

91. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. Bioinformatics. 2014 Oct 1;30(19):2811–2.

92. Wickman H, Francois R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation. 2022. Available from: https://CRAN.R-project.org/package=dplyr

93. 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 Jun 9;19(11):919–23.

94. Albert NW, Lewis DH, Zhang H, Irving LJ, Jameson PE, Davies KM. Light-induced vegetative anthocyanin pigmentation in Petunia. J Exp Bot. 2009 May 1;60(7):2191–202.

95. 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 Cell Mol Biol. 2014 Feb;77(3):367–79.

96. Butler T, Dick C, Carlson ML, Whittall JB. Transcriptome Analysis of a Petal Anthocyanin Polymorphism in the Arctic Mustard, *Parrya nudicaulis*. PLOS ONE. 2014 Jul 17;9(7):e101338.

97. 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 Apr 15;12(1):6308.

98. Sun Y, Li H, Huang JR. Arabidopsis TT19 Functions as a Carrier to Transport Anthocyanin from the Cytosol to Tonoplasts. Mol Plant. 2012 Mar 1;5(2):387–400.

99. Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, et al. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO J. 2000 Nov 15;19(22):6150–61.

100. 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 Jun 2;121(1):84–91.

101. Liu C, Yu Q, Li Z, Jin X, Xing W. Metabolic and transcriptomic analysis related to flavonoid biosynthesis during the color formation of *Michelia crassipes* tepal. Plant Physiol Biochem. 2020 Oct;155:938–51.

102. Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, et al. Advances in the MYB–bHLH–WD Repeat (MBW) Pigment Regulatory Model: Addition of a WRKY Factor and Co-option of an Anthocyanin MYB for Betalain Regulation. Plant Cell Physiol. 2017 Sep 1;58(9):1431–41.

103. Li C, Wu J, Hu KD, Wei SW, Sun HY, Hu LY, et al. *PyWRKY26* and *PybHLH3* cotargeted the *PyMYB114* promoter to regulate anthocyanin biosynthesis and transport in red-skinned pears. Hortic Res. 2020 Mar 15;7(1):1–12.

104. Shi B, Wu H, Zhu W, Zheng B, Wang S, Zhou K, et al. Genome-Wide Identification and Expression Analysis of WRKY Genes during Anthocyanin Biosynthesis in the Mango (*Mangifera indica* L.). Agriculture. 2022 Jun 7;12(6):821.

105. Xue L, Wang J, Zhao J, Zheng Y, Wang HF, Wu X, et al. Study on cyanidin metabolism in petals of pink-flowered strawberry based on transcriptome sequencing and metabolite analysis. BMC Plant Biol. 2019 Dec;19(1):423.

106. Li L, Zhai Y, Luo X, Zhang Y, Shi Q. Comparative transcriptome analyses reveal genes related to pigmentation in the petals of red and white *Primula vulgaris* cultivars. Physiol Mol Biol Plants. 2019 Jul;25(4):1029–41.

107. Xia X, Gong R, Zhang C. Integrative analysis of transcriptome and metabolome reveals flavonoid biosynthesis regulation in *Rhododendron pulchrum* petals. BMC Plant Biol. 2022 Aug 16;22(1):401.

108. Albert NW, Griffiths AG, Cousins GR, Verry IM, Williams WM. Anthocyanin leaf markings are regulated by a family of *R2R3-MYB* genes in the genus *T rifolium* . New Phytol. 2015 Jan;205(2):882–93.

109. Thill J, Miosic S, Gotame TP, Mikulic-Petkovsek M, Gosch C, Veberic R, et al. Differential expression of flavonoid 3′-hydroxylase during fruit development establishes the different B-ring hydroxylation patterns of flavonoids in Fragaria × ananassa and *Fragaria vesca*. Plant Physiol Biochem. 2013 Nov 1;72:72–8.

110. Spinardi A, Cola G, Gardana CS, Mignani I. Variation of Anthocyanin Content and Profile Throughout Fruit Development and Ripening of Highbush Blueberry Cultivars Grown at Two Different Altitudes. Front Plant Sci. 2019 Sep 4;10:1045.

111. Hong Y, Tang X, Huang H, Zhang Y, Dai S. Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in chrysanthemum. BMC Genomics. 2015 Mar 17;16(1):202.

112. Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol. 2005 Jan 1;57(2):155–71.

113. Cominelli E, Gusmaroli G, Allegra D, Galbiati M, Wade HK, Jenkins GI, et al. Expression analysis of anthocyanin regulatory genes in response to different light qualities in *Arabidopsis thaliana*. J Plant Physiol. 2008 May;165(8):886–94.

114. Araguirang GE, Richter AS. Activation of anthocyanin biosynthesis in high light - what is the initial signal? New Phytol. 2022 Dec;236(6):2037–43.

115. Ma Y, Ma X, Gao X, Wu W, Zhou B. Light Induced Regulation Pathway of Anthocyanin Biosynthesis in Plants. Int J Mol Sci. 2021 Oct 15;22(20):11116.

116. Xie G, Zou X, Liang Z, Wu D, He J, Xie K, et al. Integrated metabolomic and transcriptomic analyses reveal molecular response of anthocyanins biosynthesis in perilla to light intensity. Front Plant Sci. 2022;13.

117. Liu C, Yao X, Li G, Huang L, Xie Z. Transcriptomic profiling of purple broccoli reveals light-induced anthocyanin biosynthetic signaling and structural genes. PeerJ. 2020 May 5;8:e8870.

118. 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 Dec;20(1):110.

119. Liu C, Wang X, Shulaev V, Dixon RA. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. Nat Plants. 2016 Dec;2(12):16182.

120. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 1999 Feb 1;16(2):266–74.

121. Koes RE, Quattrocchio F, Mol JNM. The flavonoid biosynthetic pathway in plants: Function and evolution. BioEssays. 1994 Feb;16(2):123–32.

122. Lu Y, Rausher MD. Evolutionary Rate Variation in Anthocyanin Pathway Genes. Mol Biol Evol. 2003 Nov 1;20(11):1844–53.

123. 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 Mar 1;39(3).

124. Shoeva OYu, Glagoleva AYu, Khlestkina EK. The factors affecting the evolution of the anthocyanin biosynthesis pathway genes in monocot and dicot plant species. BMC Plant Biol. 2017 Dec 28;17(2):256.

125. Winkel-Shirley B. Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiol Plant. 1999 Sep;107(1):142–9.

126. Wang Y, Chen S, Yu O. Metabolic engineering of flavonoids in plants and microorganisms. Appl Microbiol Biotechnol. 2011 Aug 1;91(4):949–56.

127. Davies KM, Schwinn KE, Deroles 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 Jun 1;131(3):259–68.

128. Stafford HA. Anthocyanins and betalains: evolution of the mutually exclusive pathways. Plant Sci. 1994 Jan 1;101(2):91–8.

129. Ballester AR, Molthoff J, de Vos R, Hekkert B te L, Orzaez D, Fernaǰndez-Moreno JP, et al. Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. Plant Physiol. 2009 Dec 29;152(1):71–84.

130. 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 Sep 1;4:e2369.

131. 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 Jan;20(17):4250.

132. 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 Feb 17;12:585665.

133. 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 Jun 1;7(1):1–19.

134. 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 Nov 8;8:1917.

135. Pucker B, Walker-Hale N, Yim WC, Cushman J, Crum A, Yang Y, et al. Evolutionary blocks to anthocyanin accumulation and the loss of an anthocyanin carrier protein in betalain-pigmented Caryophyllales. 2022 Oct 21;2022.10.19.512958.

136. 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 Oct;224(1):71–85.

137. Mitsuda N, Ohme-Takagi M. Functional Analysis of Transcription Factors in Arabidopsis. Plant Cell Physiol. 2009 Jul;50(7):1232–48.

138. 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 May 7;48(8):4100–14.

139. 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 Sep 16;8(37):eabo1113.

140. Latchman DS. Transcription factors: An overview. Int J Biochem Cell Biol. 1997 Dec;29(12):1305–12.

141. 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 Dec;175(7):1842-1855.e16.

142. 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 Oct 10;52(1):25–36.

143. 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 Feb;22(2):96–118.

144. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 2001 Oct 1;4(5):447–56.

145. 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.

146. Pireyre M, Burow M. Regulation of MYB and bHLH Transcription Factors: A Glance at the Protein Level. Mol Plant. 2015 Mar 2;8(3):378–88.