Genetic factors explaining anthocyanin pigmentation differences

Maria F. 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 genetic factors causing anthocyanin loss were reported. Additionally, transcriptomic data sets from previous studies were reanalyzed 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. Our findings underline the particular significance of MYBs and their apparent dominant role in the specificity of the MBW complex.

Keywords: Flavonoid biosynthesis; Anthocyanins; Gene expression; Transcription factor; MYB; Systematic comparison; Transcriptome.

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 show a constant transformation 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].

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 ranging 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]. Several reports suggest that the interaction with copigments like flavonols and flavones is

an important factor for the stabilization of anthocyanins in plants[17,18]. 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[19], 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 functions and can influence the coloration of different plant structures[20]. 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"[21], *Chrysanthemum grandiflorum* cv. Jinba[22], and *Chrysanthemum morifolium*[23]. Proanthocyanidins also called condensed tannins are colorless compounds that turn brown upon oxidation[24]. They have been studied in seeds of species such as *Arabidopsis thaliana*[25], *Brassica napus*[26], and *Ipomoea purpurea*[27].

The general pathway of the flavonoid biosynthesis (Figure 1A) is well understood and the central aglycon biosynthesis is largely conserved among land plants[3,28]. 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 oxidated 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 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. This step also requires an anthocyanin-related gluthationine S-transferase (arGST) that was named AN9/TT19 due to the corresponding mutants[29,30], but the enzymatic function was only revealed recently [31]. 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.

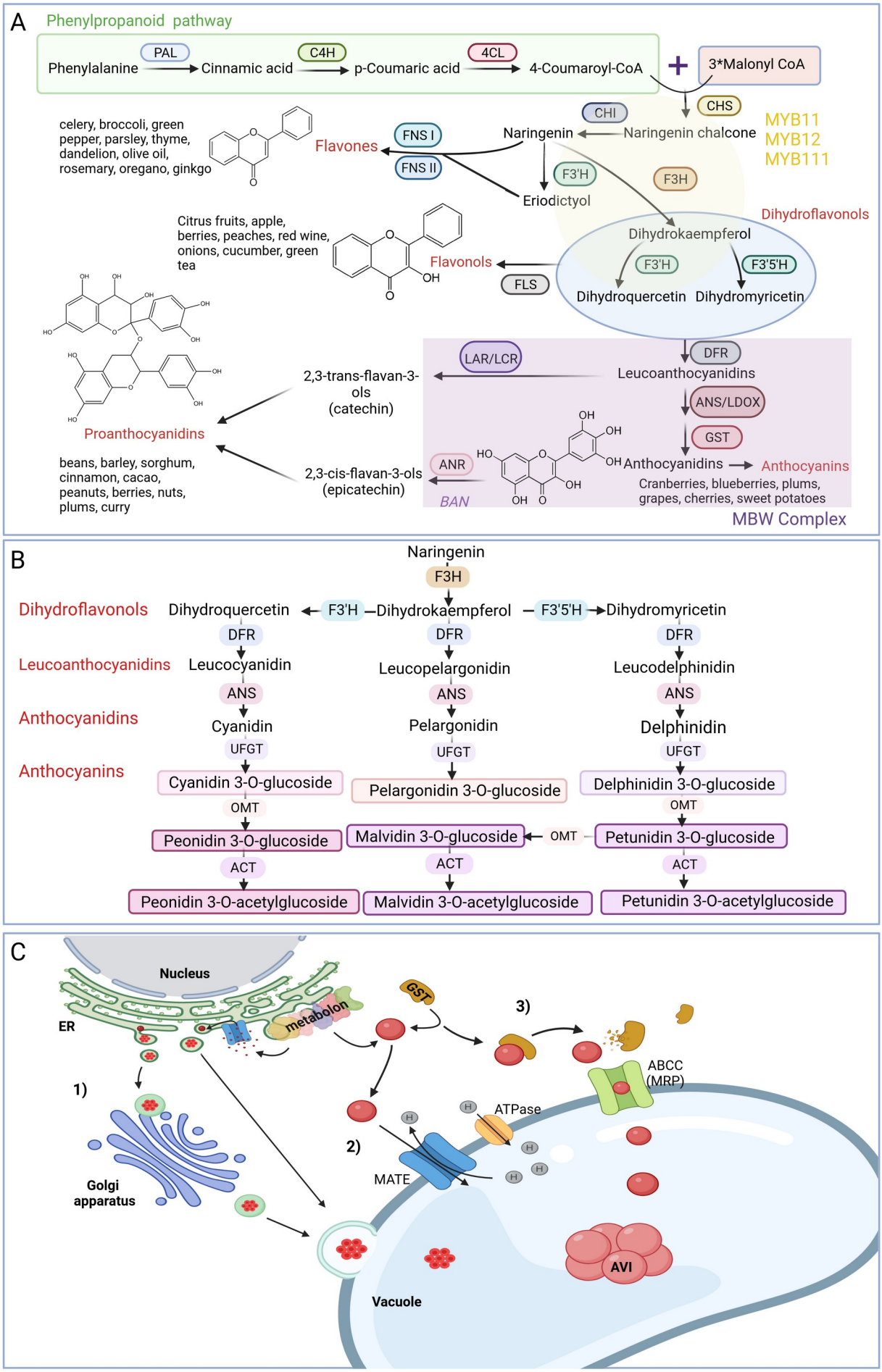
A
Phenylpropanoid pathway
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
4-Coumaroyl-CoA
3*Malonyl CoA
CHS
CHI
Naringenin chalcone
Naringenin
MYB11
MYB12
MYB111
celery, broccoli, green pepper, parsley, thyme, dandelion, olive oil, rosemary, oregano, ginkgo
Flavones
FNS I
FNS II
F3'H
F3H
Eriodictyol
Dihydroflavonols
Dihydrokaempferol
Citrus fruits, apple, berries, peaches, red wine, onions, cucumber, green tea
Flavonols
FLS
F3'H
F3'5'H
Dihydroquercetin
Dihydromyricetin
DFR
LAR/LCR
Leucoanthocyanidins
2,3-trans-flavan-3-ols (catechin)
ANS/LDOX
GST
Proanthocyanidins
ANR
Anthocyanidins
Anthocyanins
beans, barley, sorghum, cinnamon, cacao, peanuts, berries, nuts, plums, curry
2,3-cis-flavan-3-ols (epicatechin)
BAN
Cranberries, blueberries, plums, grapes, cherries, sweet potatoes
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
1)
2)
3)
ABCC (MRP)
ATPase
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, ANS- Anthocyanidin syntase, LDOX - leucoanthocyanidin dioxygenase, FNS - flavone synthase. (C) Scheme of the different intracellular flavonoid transport mechanisms. 1. vesicle trafficking from the Endoplasmic Reticulum (ER) 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.*[31]. It is currrently not clear if GST functions only as enzyme or if they also play a role in anthocyanin transport.

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 endoplasmic reticulum (ER) and form a metabolon[32](Figure 1C). Flavonoids produced at the ER are transported into the vacuole for storage[33–35] 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[34]. 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[32,35]. The existence of these vesicles has been reported mainly in *Zea mays*[36], *A. thaliana*[37], *Vitis vinifera*[38], and *Oryza sativa*[39]. Microscopic evidence shows that

these vesicular bodies are attached to the surface of the ER[38,40] 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[41]. The GST-mediated transport model proposes that flavonoids are delivered to the tonoplast by ligandins[29,42]. These 'ligandins' would be glutathione S-transferase (GST) binding and carrier proteins[43]. Evidence for the role of GST in flavonoid transport has been reported in multiple species, such as *Zea mays*[42], *A. thaliana*[29,30], *Petunia hybrida*[43], *Vitis venifera*[38], and *Prunus persica*[44]. Reconsideration of the GST-mediated transporter is needed in the light of a recent study[31] 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[32]. 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[45]. 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[45–47].

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[48–50]. The mechanism that regulates the expression of all structural genes in the anthocyanin biosynthesis pathway is commonly known as the MBW complex[51]. 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 combinatoric diversity[51]. After the discovery of TT8 in *A. thaliana*[52], Baudry *et al*.[53] 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[54]. 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[54].

It is assumed that the function of regulating anthocyanin biosynthesis by the MBW complex is evolutionary conserved across angiosperms[55]. Elements of the MBW complex are also involved in the control of other biological processes in plants[56]. Other functions add different degrees of evolutionary constraints to the components of the MBW complex. MYB partners are substituted and considered as the specificity determining factor in the MBW complex. For example, PAP1/MYB75 and PAP2/MYB90 activate the anthocyanin biosynthesis[57], while TT2/MYB123 would activate the proanthocyanidin biosynthesis[53]. The DNA binding and protein-protein interaction capacity of MYBs and bHLHs is determined by highly conserved regions[58]. It has been postulated that TTG1 serves as a scaffolding protein that maintains the interaction of MYB and bHLH[59]. 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[60–62].

Closely related plant species can differ in their anthocyanin repertoire and pigmentation pattern[63,64]. These differences can even appear between plants of the same species[1,65–67]. 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*[68], *Vitis vinifera*[69], *Hordeum vulgare*[70], *Nicotiana tabacum*[71], and *Nicotiana alata*[72]. Substrate competition between different branches of the flavonoid biosynthesis can also have an impact on the anthocyanin accumulation[73]. For example, an increased flavonol production can lead to a pigmentation loss due to insufficient substrate for the anthocyanin biosynthesis[74]. 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[75], but previous research suggests that some genes are more often responsible for pigmentation loss than others[62,76]

This raises the question whether there are “hot spots” in the pathway that are frequently the cause for a pigmentation loss; not due to higher mutation rate, but higher likelihood of mutation fixation. Specifically,

we ask three questions: (1) Are anthocyanin pigmentation differences within species caused mainly by transcription factors or structural genes? (2) Is *DFR* at the start of the anthocyanin biosynthesis branch more likely to be causal for anthocyanin variations than downstream genes like *ANS*, *arGST*, or *UFGT*? (3) Are MYB more often responsible for pigmentation differences than other transcription factor families? To address these questions, we performed a systematic comparison of anthocyanin-pigmented vs. non-pigmented varieties across flowering plant species. The literature was screened for reports of causal genes explaining pigmentation differences between varieties of a plant species or between closely related (sub)species. Studies without conclusive results were checked for data availability and subjected to reanalysis when possible. The contribution of different genes to the pigmentation differences was quantified and supported a crucial role of transcription factors and especially MYBs.

RESULTS

Genetic hot spots responsible for anthocyanin pigmentation differences

Based on a total of 230 analyzed studies (see Additional file1: 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. In our schematic representation (Figure 2) , we defined pigmentation to be the wild type state, while absence of pigmentation was defined to be the result of a mutation. We identified 13 research articles that report upregulated structural genes in pathways competing for substrate with the anthocyanin biosynthesis as the cause of color difference between unpigmented and pigmented accessions. 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 likely 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.

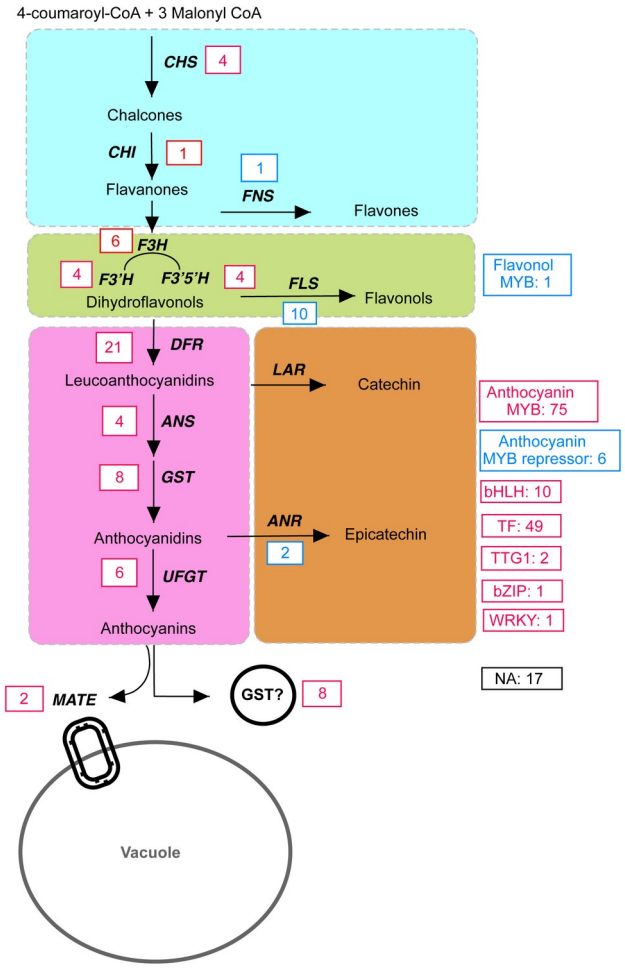

Figure 2. Schematic representation of the flavonoid biosynthesis pathway with emphasis on the number of cases in which a particular gene was responsible for color difference according to a systematic literature analysis and re-analyses of RNA-seq data sets. The anthocyanin-pigmented accession is set as reference when determining up- and down-regulation. Blue boxes and red boxes indicate the number of upregulated and downregulated/non-functional/lost genes, respectively. Downregulated 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 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, TF – unclassified transcription factor. GST is presented twice due to the recently reported enzymatic function of

arGSTs by Eichenberger *et al.* [31]. It is currrently not clear if GST functions only as enzyme or if they also play a role in anthocyanin transport.

Taxonomic distribution of analyzed cases across plant families

The 230 reviewed studies (see Additional file 1: Table S1), cover 53 plant families distributed over 30 orders (Figure 3). 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 each in 7 cases. Other families, including Lamiaceae, Asparagaceae, Malvaceae and Caryophyllaceae, were identified in varying frequencies across the reviewed studies. 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 when comparing intraspecific plants. 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. This nuance would be helpful for the interpretation of the results to distinguish between naturally fixed variants and those selected by breeders.

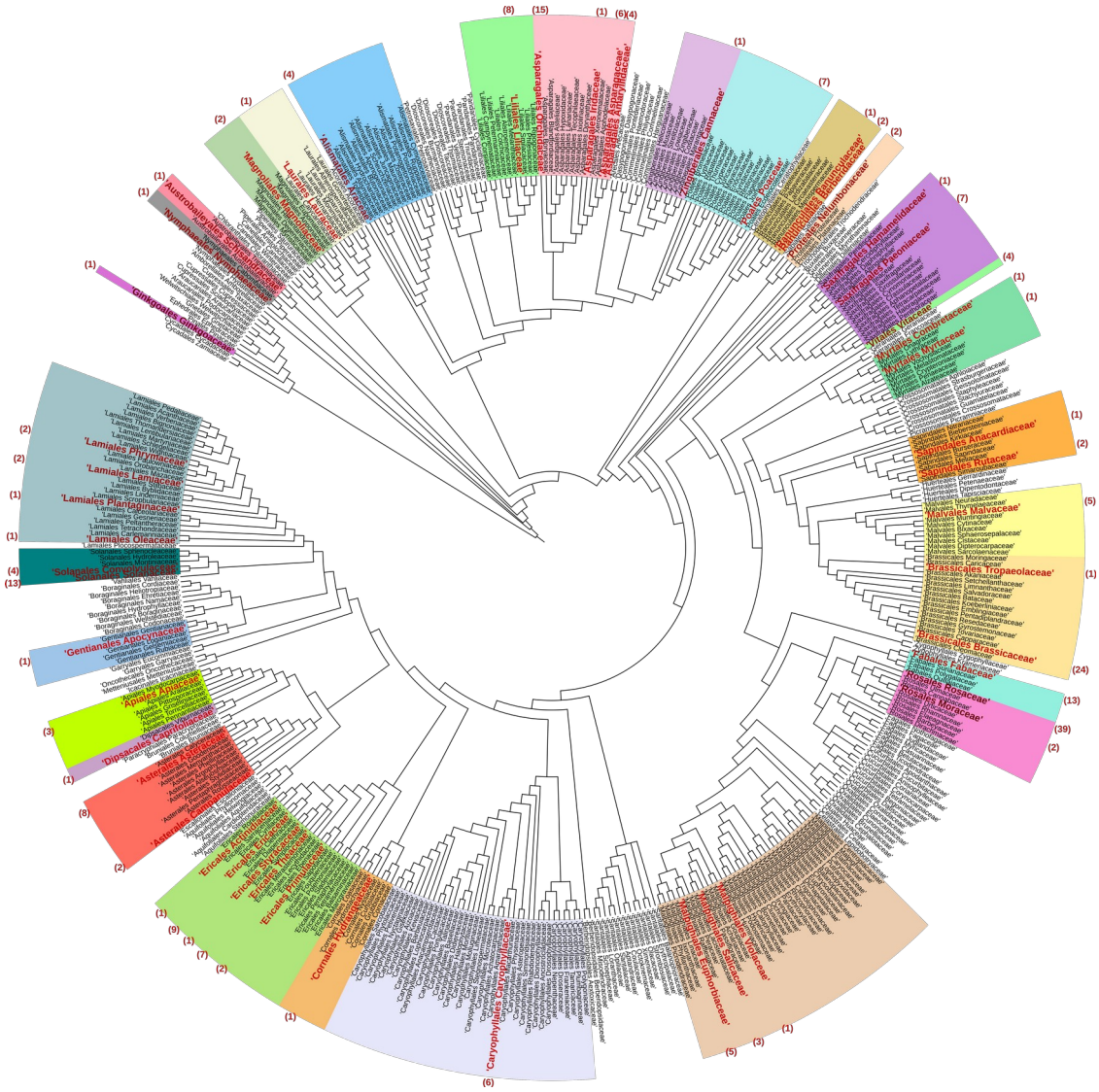

Figure 3. Phylogenetic tree displaying 428 angiosperm families. Each color range groups the families of an order. Families highlighted in bold red are those encompassed in the literature screening for anthocyanin pigmentation differences. The number of pigmentation difference cases is given in parentheses for each family. Modified from Li *et al.[77]*.

Genetic factor of anthocyanin loss across plant families

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 (Figure 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 particular 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 causal for anthocyanin loss with the respective families where they have been observed, we aim to discern any family-specific patterns.

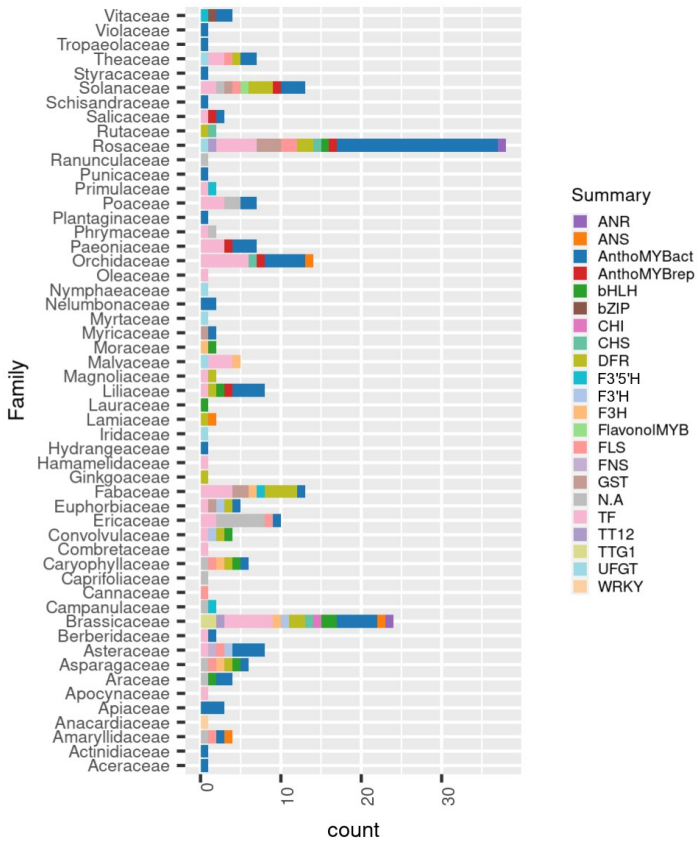

Figure 4. Number of cases each gene is implicated in anthocyanin differences reported in the literature (Additional file 1: Table S1) resolved by family. ANR - Anthocyanidin reductase, ANS - Anthocyanidin synthase, AnthoMYBact – Anthocyanin MYB activator, AnthoMYBrep – Anthocyanin MYB repressor, bHLH – 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, TF – Transcription factor, TT12 – Transparent Testa 12, TTG1- Transparent Testa

Glabra 1, UFGT - UDP-glucose:flavonoid 3-O-glucosyltransferase, WRKY- WRKY DNA-binding domain, N.A – Unknown factor.

Overrepresentation of genetic factors causing anthocyanin loss

The hypothesis that *DFR* might be more often responsible for an anthocyanin loss than *ANS* or other players in the anthocyanin biosynthesis was tested. Among the 230 studies analyzed (see Additional file 1: Table S1), *DFR* exhibited the highest frequency, being reported in 21 cases, while *ANS* was identified in four cases as the gene responsible for 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, p-value = 0.00067).

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, p-value = 0.0158). Similarly, when examining the relationship between *DFR* and *UFGT*, a notable difference was again detected (χ^2 test, p-value = 0.0039). This suggests that *DFR* is the most important target of evolutionary events blocking the anthocyanin pathway.

To investigate if the color variation could be attributed to the substrate competition between FLS and DFR, we examined the prevalence of cases revealing *FLS* and *DFR* as the primary genes influencing color variation in plant tissue. A hyperactivation of *FLS* was identified in 10 cases as the factor responsible for pigmentation differences which is significantly lower than the number of 21 cases in which a DFR down-regulation/loss was responsible (χ^2 test, 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 downregulation 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.

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 significant higher than the presence of other transcription factor families such as bHLH, WRKY, TTG1, bZIP, and others (χ^2 test, p-value = 1.61e-10). Even when compared against the large group of unclassified TFs, we observed that the presence of MYBs is significantly higher (χ^2 test, p-value = 0.019). It is expected that most of the cases with unidentified transcription factors would actually be due to anthocyanin biosynthesis activating MYBs. In summary, these findings collectively highlight the dominant 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 e.g. to indicate which flowers have already been visited by insects[78]. Flavonoids are also known to protect against UV radiation[79], drought[80], and cold stress[81,82]. 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 candidates genes responsible for color variation. This analysis harnessed numerous public RNA-seq data sets that enable a direct comparison of anthocyanin biosynthesis gene activity between differently pigmented accessions and the taxonomical family they belonged. 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[83,73,84–87] and has been experimentally tested in families such as Solanaceae[86], as well as in specific genera like Ipomoea[61,63], Iris[88], Antirrhinum[89], and Petunia[76].

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 harbouring 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[90]. 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[91].

A study published by Rausher *et al*.[92] proposed that the evolutionary rate of enzymes depends on their location in a pathway with early genes showing a slower mutation rate, but this has been contradicted recently[93]. 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*. 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[94,95]. 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[17]. Metabolically, this can occur when two enzymes or transporters compete for the same or very similar substrates[94]. 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, 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[96]. A recent study identified a flavonol biosynthesis regulating MYB as the most frequently affected gene in pigmentation pattern change[93]. 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 sets 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[97–100] and bHLHs[101–103] 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[95]. Similarly, Wheeler *et al*.[93] showed that transcription factors, particularly MYB, 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 Petunieae tribe[93]. This premise commonly known as the E-R anticorrelation, has been widly studied across different organisms, including yeast[104,105], Arabidopsis[106], Brassica[107], Barley[108], Arachis[109], and Drosophila[110]. 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 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[111], which do not accumulate anthocyanins[96,112].

Studies on evolutionary rates investigating the components of the MBW complex suggested that MYBs would be the most likely component to be lost due to the highest degree of specialization which coincides with a lower pleiotropy[62,113,93]. It was observed that insertions/deletions were the most frequent mutation events in MYBs, while amino acid substitutions in the conserved region appeared irrelevant[61]. This let 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[61]. 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[111]. 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 [13,114,115]. The lack of a functional MBW complex is considered as one crucial factor for the loss of

anthocyanin pigmentation in the betalain-pigmented Caryophyllales[111]. 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 likely to accumulate mutations than non-TF genes[116]. Another study examined the frequency of point mutations in TF 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[117]. 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[93]. 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*[118]. 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[118].

The number of TFs required to regulate a specific enzyme-encoding gene is a complex and dynamic process that is influenced by various factors[119], including the complexity of the regulatory region[120], the stage of development or cell type[121], and environmental factors[122]. For example, only about ten *R2R3-MYBs* are known to bind specific DNA motifs related to the regulation of flavonoid biosynthesis pathway in *A. thaliana*[50,123]. In a review study, Feller *et al.*[58] 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[58]. Multiple bHLH proteins contain a similar ligand-binding domain targeting different enzyme encoding genes[124]. A greater proportion of MYBs were recognized to be responsible for the regulation of flavonoid biosynthesis genes in comparison to bHLHs[124]. 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 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.

CONCLUSION

A 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. The MYB 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 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.

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 evidences 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*[125], *Rhododendron obtusum*[126], *Trifolium repens*[127], and *Hosta plantaginea*[128] (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[129].

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[130] 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[131] to enable the following assembly with Trinity. Trinity v2.4.0[132] 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[133] (Additional file 1: Table S3).

After completion of the assembly process, kallisto v0.44[134] 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 constructed 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[135] with the package ggplot2 v.3.4.0[136] to inspect the variation within the data set (Additional file 3: Figures S1-S4).

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[137] that combines Transdecoder[138], ORFfinder[139], and ORFpredictor[140]. Knowledge-based Identification of Pathway Enzymes (KIPEs3) v0.34[141,142] 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[143]. A previously described BLAST-based Python script[144] was deployed for the identification of additional candidates genes using bHLH, WD40, and WRKY genes associated with the flavonoid biosynthesis as baits[131]. 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[145] 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[146] using default parameters and RAxML v.8.2.12[147] with PROTGAMMA+LG+F. Phylogenetic trees for the transcription factor families MYB, bHLH, TTG1, and WRKY were constructed to assess orthologues relationships (Additional file 2: 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.[77]* and modified with iTOL v.6.8.1[148] 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[149], circlize 0.4.15[150] and dplyr 1.1.0[151]. 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[131]. A workflow of the methods used in the transcriptome analysis is described in Figure 6.

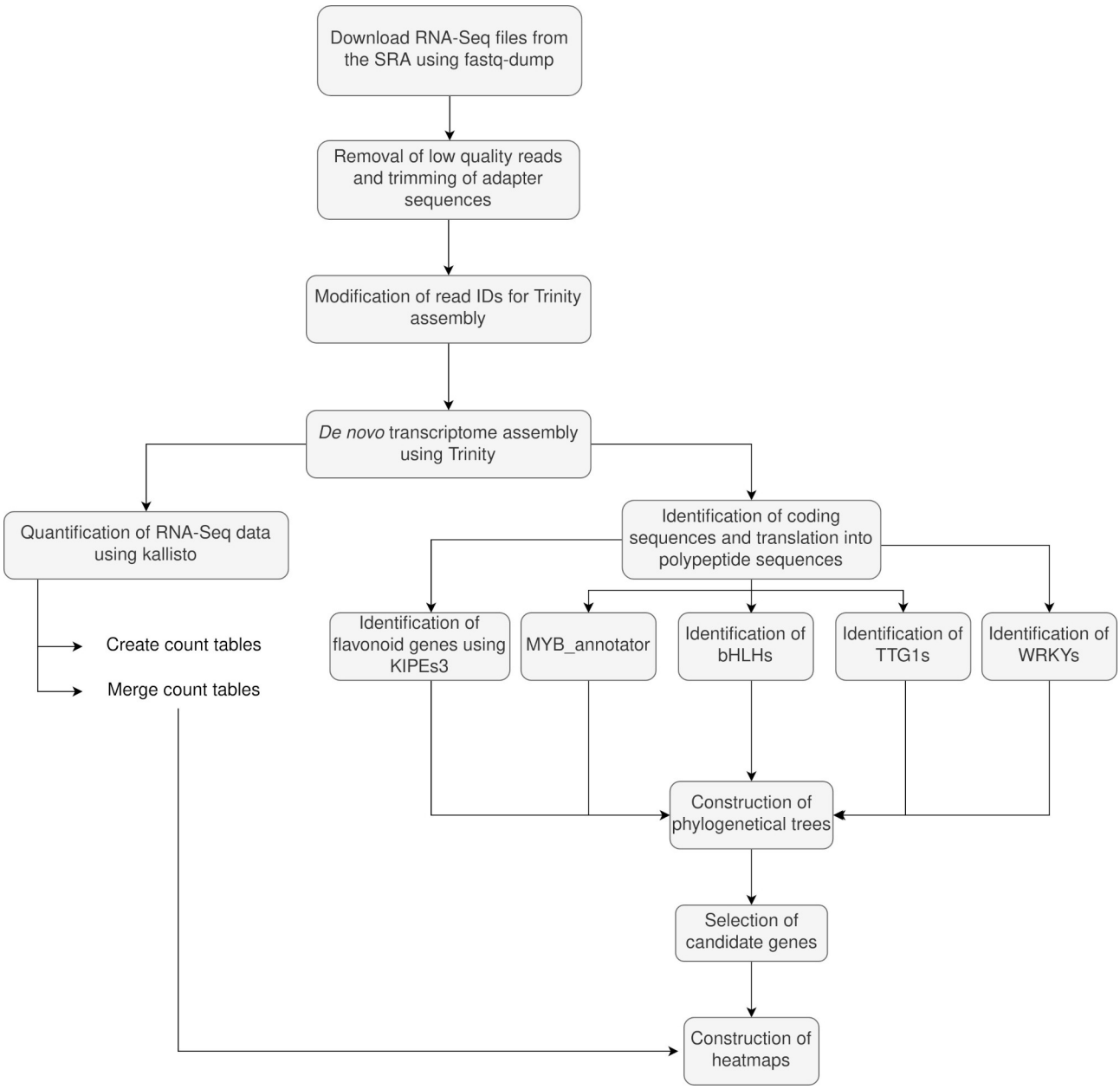

Figure 6. Flowchart representation of the methodology followed in the comparative transcriptional analysis. SRA – Sequence Read Archive, KIPEs – Knowledge-based Identification of Pathway Enzymes, MYB - Myeloblastosis, bHLH – basic helix-loop-helix, *TTG1 – Transparent Testa Glabra 1*, WRKY – WRKY DNA-binding domain.

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 Glabra 1*;* WRKY: WRKY DNA-binding domain.

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. All the tests were applied using the statistical software R v.4.1.3[135].

Declarations

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. We used bioRender.com for the construction of some figures.

Authors' contributions

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

Funding

This work received no external funding.

Availability of data and materials

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).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

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. Ruiz-Sola MÁ, Rodríguez-Concepción M. Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. The Arabidopsis Book. The American Society of Plant Biologists. 2012(10).

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, de Souza LP, Fernie AR. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. J Exp Bot. 2017 Sep 9;68(15):4013–28.

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. 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.

18. Mazza G, Brouillard R. The mechanism of co-pigmentation of anthocyanins in aqueous solutions. Phytochemistry. 1990 Jan;29(4):1097–102.

19. Horbowicz M, Kosson R, Grzesiuk A, Dębski H. Anthocyanins of Fruits and Vegetables - Their Occurrence, Analysis and Role in Human Nutrition. J Fruit Ornam Plant Res. 2008 Jan 1;68(1):5–22.

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

21. 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.

22. Zhang Z. The identification of flavonoids and the expression of genes of anthocyanin biosynthesis in the chrysanthemum flowers. :7.

23. 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.

24. 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.

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

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

27. 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.

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

29. 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.

30. 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.

31. 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 Aug 31;1–12.

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

33. 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.

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

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

36. 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.

37. 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.

38. 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.

39. 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 2;10(4):685.

40. 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.

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

42. Marrs KA, Alfenlto MR, Lloyd AM. A glutathione $-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. 1995;375.

43. 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;10(7):1135–49.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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 May;50(4):660–77.

50. 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.

51. Ramsay NA, Glover BJ. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 2005 Feb 1;10(2):63–70.

52. 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.

53. 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.

54. 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.

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

56. 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).

57. Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell. 2000 Dec;12(12):2383–94.

58. 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.

59. 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 Jun 1;8(3):272–9.

60. Martin C, Ellis N, Rook F. Do Transcription Factors Play Special Roles in Adaptive Variation? Plant Physiol. 2010 Oct 1;154(2):506–11.

61. Streisfeld MA, Liu D, Rausher MD. Predictable patterns of constraint among anthocyanin-regulating transcription factors in Ipomoea. New Phytol. 2011;191(1):264–74.

62. Streisfeld MA, Rausher MD. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution: mutations during adaptive evolution. Evolution. 2011 Mar;65(3):629–42.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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 Jan 24;7(1):12.

70. 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.

71. 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 Sep 7;21(1):611.

72. 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;12(12):1976.

73. Rausher MD. Evolutionary Transitions in Floral Color. Int J Plant Sci. 2008 Jan;169(1):7–21.

74. 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.

75. 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.

76. Wheeler LC, Dunbar-Wallis A, Schutz K, Smith SD. Evolutionary walks through flower colour space driven by gene expression in Petunia and allies (Petunieae). Proc R Soc B Biol Sci. 2023 Jul 5;290(2002):20230275.

77. 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 Oct 29;19(1):232.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE. Regulatory Genes Controlling Anthocyanin Pigmentation Are Functionally Conserved among Plant Species and Have Distinct Sets of Target Genes. :17.

84. Sobel JM, Streisfeld MA. Flower color as a model system for studies of plant evo-devo. Front Plant Sci. 2013.

85. 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.

86. Ho WW, Smith SD. Molecular evolution of anthocyanin pigmentation genes following losses of flower color. BMC Evol Biol. 2016 May 10;16(1):98.

87. Sapir Y, Gallagher MK, Senden E. What Maintains Flower Colour Variation within Populations? Trends Ecol Evol. 2021 Jun 1;36(6):507–19.

88. Roguz K, Gallagher MK, Senden E, Bar-Lev Y, Lebel M, Heliczer 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.

89. Goodrich J, Carpenter R, Coen ES. A common gene regulates pigmentation pattern in diverse plant species. Cell. 1992 Mar 6;68(5):955–64.

90. 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.

91. 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.

92. 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.

93. 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).

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

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

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

97. 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. 2010 Jan 1;152(1):71–84.

98. 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.

99. 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.

100. 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.

101. 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.

102. 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.

103. 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.

104. Pál C, Papp B, Hurst LD. Highly expressed genes in yeast evolve slowly. Genetics. 2001 Jun;158(2):927–31.

105. 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 May 13;6(5):e1000944.

106. Yang L, Gaut BS. Factors that Contribute to Variation in Evolutionary Rate among Arabidopsis Genes. Mol Biol Evol. 2011 Aug 1;28(8):2359–69.

107. 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.

108. 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

109. 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 Nov 1;7(1):14853.

110. 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 Jul 1;21(7):1308–17.

111. Pucker B, Walker-Hale N, Dzurlic J, Yim WC, Cushman JC, Crum A, et al. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented Caryophyllales, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. New Phytol.2023.

112. 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.

113. Guillaume F, Otto SP. Gene Functional Trade-Offs and the Evolution of Pleiotropy. Genetics. 2012 Dec 1;192(4):1389–409.

114. 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.

115. 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 Jan;47(1):92–6.

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

117. 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.

118. 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.

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

120. 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.

121. 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.

122. 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.

123. 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.

124. 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.

125. 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:e4393905.

126. 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 1;49(4):2641–53.

127. 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 1;110(3):191–200.

128. 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.

129. 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.

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

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

132. 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.

133. 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).

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

135. 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/.

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

137. 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.

138. 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.

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

140. 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.

141. 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.

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

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

144. Pucker B, Iorizzo M. Apiaceae FNS I originated from F3H through tandem gene duplication. PLOS ONE. 2023 Jan 19;18(1):e0280155.

145. 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.

146. 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.

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

148. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.

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

150. 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.

151. 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