Conserved amino acid residues and gene expression patterns associated with the substrate preferences of the competing enzymes FLS and DFR

Nancy Choudhary[1] and Boas Pucker[1,*]

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

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

Abstract

Flavonoids, an important class of specialized metabolites, are synthesized from phenylalanine and present in almost all plant species. Different branches of flavonoid biosynthesis lead to products like flavones, flavonols, anthocyanins, and proanthocyanidins. Dihydroflavonols form the branching point towards the production of non-coloured flavonols via flavonol synthase (FLS) and coloured anthocyanins via dihydroflavonol 4-reductase (DFR). An angiosperm-wide comparison of FLS and DFR sequences revealed the substrate specifying amino acids in both enzymes. We identified the patterns of these residue occurrences in angiosperms and revealed that monocots generally possess FLS with Y132 residue (FLS_Y) and DFR with N133 residue (DFR_N). In contrast, dicots generally possess FLS_H and DFR_N, DFR_D and DFR_A. Generally, angiosperm FLS prefers dihydrokaempferol, whereas DFR favors dihydroquercetin or dihydromyricetin. This study illustrates the importance of substrate-specifying residues in resolving competition between the FLS and DFR enzymes, delineates the evolution of these enzymes, and provides insights into mechanisms directing the metabolic flux of the flavonoid

biosynthesis, with potential implications for ornamental plants and molecular breeding strategies.

Keywords: Flavonoid Biosynthesis, Substrate competition, FLS vs DFR, Substrate specificity

Background

Flavonoids comprise a range of different subclasses including flavones, flavonols, anthocyanins and proanthocyanidins, all of which have their own biological functions (1). Flavonoids exhibit a remarkable diversity, encompassing a vast array of compounds that possess distinct properties, conferring specific evolutionary benefits (2). The biosynthesis of these compounds involves a complex network of enzymatic reactions. Mutants with observable phenotypic variations in flavonoid biosynthesis have served as valuable model systems for studying the intricacies of biosynthetic pathways and their regulatory mechanisms (3–6).

Anthocyanins are renowned for their role as red, purple, or blue pigments, contributing to the vibrant colors seen in flowers and fruits (1). These color patterns serve to attract animals for important biological processes such as pollination and seed dispersal (7). Flavonols have been described as co-pigments that contribute to the overall coloration of various plant structures (8,9). On their own, flavonols are colorless or pale yellow and have been described as important factors in the UV stress response (10). Additionally, the flavonoid biosynthesis pathway produces other compounds such as flavones and proanthocyanidins (11,12). Flavones play a crucial role in signalling and pathogen defense (11), whereas proanthocyanidins are key contributors to seed coat pigmentation (13). Beyond their pigmentation functions, flavonoids have diverse roles in response to environmental factors like drought (14), cold (15–17), and highlight (18,19), playing a vital physiological role in protecting against UV-induced oxidative stress (20,21), detoxifying reactive oxygen species (ROS) (22,23) and toxic metals (24),

defending against pathogens (25), and regulating symbiotic associations (26,27) and influencing auxin flux (28,29).

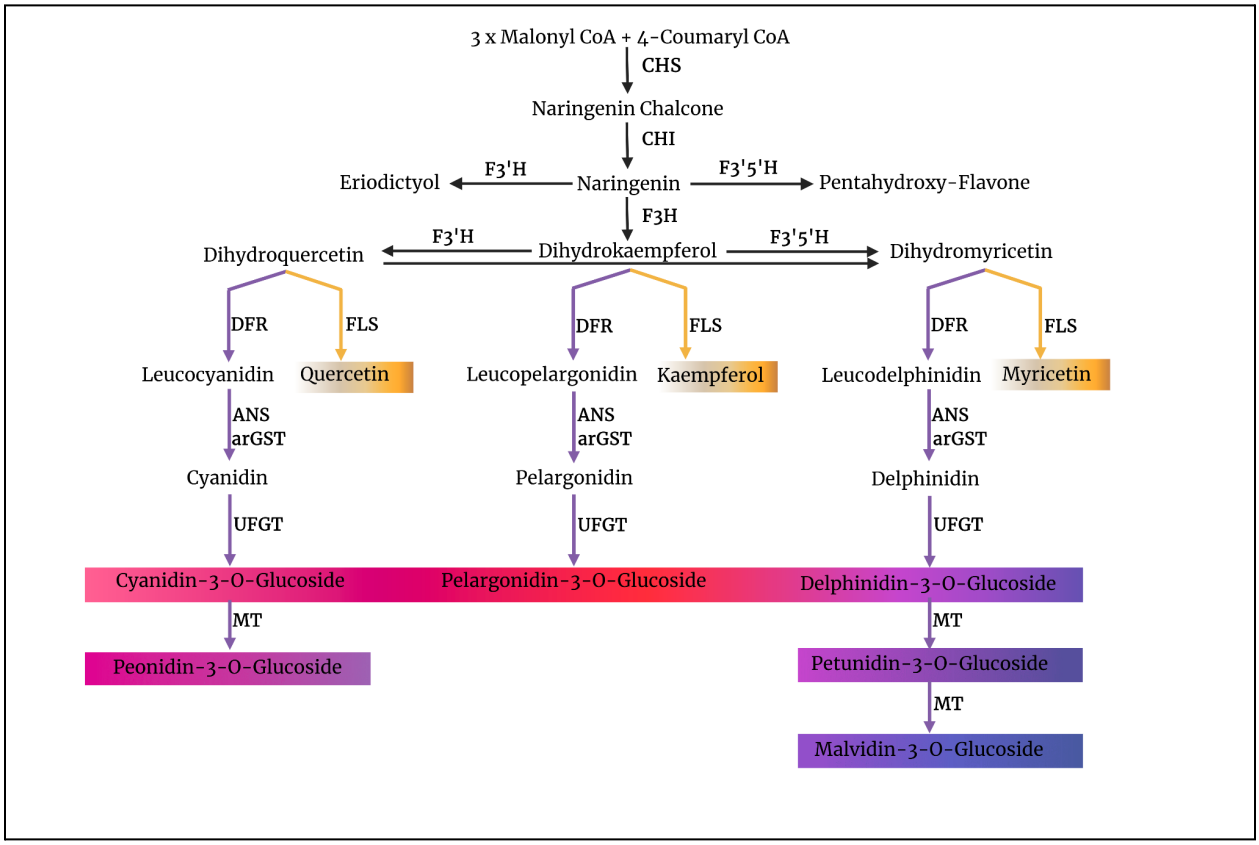

Figure 1: Simplified Flavonoid Biosynthesis Pathway in plants showing the competition between FLS and DFR for common substrate, dihydroflavonols. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; FLS, fllavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; arGST, anthocyanin-related glutathione S-transferase; UFGT, UDP-glucose flavonoid 3-O-glucosyl transferase; MT, methyltransferase.

A range of different colors is provided by three classes of anthocyanins: pelargonin (orange to bright red), cyanin (red to pink), and delphinin (purple or blue) (1). These three classes are characterized by different numbers of hydroxyl groups at the B phenyl ring: pelargonin (one),

cyanin (two), and delphinin (three) (30). The biosynthesis of anthocyanins involves the enzymatic conversion of colorless dihydroflavonols into leucoanthocyanidins by DFR (31). ANS further converts leucoanthocyanidins into anthocyanidins. Recent research suggests the involvement of a glutathione-S transferase in anthocyanin biosynthesis (32,33). Anthocyanidins are transformed into anthocyanins through the addition of a sugar moiety by glycosyl transferases (34). Subsequent modifications such as the addition of other sugar moieties (35) or acyl groups (36) are catalyzed by a large number of often promiscuous enzymes.

Dihydroflavonols are not exclusively channeled into anthocyanin biosynthesis but are also the substrate of the flavonol synthase (FLS) that produces kaempferol, quercetin, and myricetin depending on the available dihydroflavonol (37). This leads to a competition for the dihydroflavonol substrates between DFR and FLS (Figure 1). Previous studies suggest that the ratio of FLS to DFR activity is an important factor in color determination (37–41). A study across multiple flowering plants discovered that high FLS activity can result in a lack of anthocyanin pigmentation despite a functional anthocyanin biosynthesis machinery being present (40). The flavonoid 3'-hydroxylase (F3'H) converts dihydrokaempferol into dihydroquercetin by the addition of one hydroxyl group in the C3' position of the B ring, while the flavonoid 3',5'-hydroxylase (F3'5'H) converts dihydrokaempferol into dihydromyricetin by adding hydroxyl groups at the C3' and C5' positions (42). Differences in the stability of the different dihydroflavonols pose a challenge for quantitative *in vitro* studies (43).

The activity of different branches of the flavonoid biosynthesis is controlled by transcriptional activation of the involved genes. *FLS* gene expression is activated by MYB transcription factors of the subgroup7 and this mechanism appears to be generally conserved across plant species with minor lineage-specific variations (44,45). Recently, a pollen-specific activation of *FLS* by sugroup19 MYBs was described (46,47). Transcriptional activation of DFR is controlled by an

MBW transcription factor complex that is named after the three components: MYB, bHLH, and WD40 (48). While research on *A. thaliana* mutants suggests that positive regulation dominates the gene expression control of flavonoid biosynthesis genes, several negative regulators have been identified through research on many different plant species (49).

Apart from the overall activity of FLS and DFR, the substrate selectivity plays a crucial role in determining the production of flavonols and anthocyanins (50,51). This selectivity is evident in species like *Freesia hybrida*, in which the presence of abundant kaempferol glycosides and delphinidin derivatives (52) indicates the separation of flavonol and anthocyanin biosynthesis based on the contrasting substrate preferences of DFR and FLS (52). Despite testing various DFR copies in *Freesia hybrida*, none of them exhibited a preference for DHK, while all favored DHM as the substrate. Some DFRs have been found to accept all three potential dihydroflavonols, indicating that the availability of specific substrates determines the class of anthocyanins produced. Other DFRs are incapable of accepting dihydrokaempferol as a substrate, resulting in the absence of orange anthocyanin-based pigmentation in species like petunia, apple, pear, hawthorn, rose, and cranberry (43,51,53). The substrate specificity determining region was predicted by a multiple alignment of polypeptide sequences (54). This region was further investigated based on the *Gerbera hybrida* DFR which is able to accept DHK and leads to orange flower pigmentation (55). Two important amino acid residues implicated in the substrate specificity of DFR were identified based on chimeric *Gerbera hybrida* and *Petunia hybrida* DFRs followed by site-directed mutation experiments (51). The relevance of all positions in a region of 26 amino acid residues was evaluated by permutation experiments. The mutation E145L (of the *G. hybrida* DFR resulted in a non-functional enzyme while the N134L mutation restricted the substrate preference to dihydrokaempferol and excluded dihydroquercetin or dihydromyrecitin (51). Many following studies confirmed the particular relevance of this N134 residue (position 133 in *Arabidopsis thaliana* DFR) and identified additional amino acid residues

with specific properties (50,56–58). In summary, N allows the acceptance of all dihydroflavonols, L restricts the substrate to DHK, D enables acceptance of DHQ and DHM and A leads to a high DHK affinity.

While the substrate preference of DFR received substantial attention, little is known about the substrate preferences of FLS. The variable preferences of flavonol synthases for 3'-hydroxylated and 3',4'-hydroxylated substrates have been investigated. Studies highlighted that some flavonol synthases, such as *Oryza sativa OsFLS* (59)*, Zea mays ZmFLS1* (60)*, Arabidopsis thaliana AtFLS1* (61), *Citrus unshiu CitFLS* (62), and *Camellia sinensis CsFLSs* (63) exhibit greater efficiency in using 3'-hydroxylated dihydrokaempferol (DHK) over 3',4'-hydroxylated dihydroquercetin (DHQ), while *Allium cepa AcFLS* (64), *Fagopyrum tataricum FtFLS1* (65), *Ipomoea batatas IbFLS* (66), and *Nicotiana tabacum NtFLS* (67) show the opposite trend. Five specific residues, namely H132, F134, K202, F293, and E295, have been reported to be involved in the binding of the DHQ substrate (68). Previous research in *A. thaliana* was established through the incubation of *AtFLS* mutants with DHQ. The mutants, H132F and H132Y exhibited 124% and 83% activity, respectively, compared to the wild-type (WT). The K_M values of mutants H132F and H132Y displayed an approximately 2-fold decrease compared to the WT counterpart (68). Conversely, mutations in the other four residues led to a significant decrease in FLS activity (68). Interestingly, the four highly conserved residues were found to be crucial for substrate binding and catalytic activity. The H132 residue was found to be variable across different flavonol synthases, with some having F residues and others having H residues. These findings suggested that the H132 position may play a role in determining substrate preferences among different flavonol synthases (68). It is postulated that the position of H132 interacts via a hydrogen bond with the B-ring hydroxyl group of the dihydroflavonol substrates (69).

While the model organism *A. thaliana* has only a single DFR gene (70), multiple DFR copies have been reported in many other plant species. For example, studies in buckwheat (71) and strawberry (72) have identified DFR copies that show substantial differences in their substrate preference and consequently also in the produced anthocyanin classes. Genetic analysis of *DFR1* and *DFR2* in buckwheat suggested that both genes are genetically linked, but the genetic distance does not suggest a recent tandem duplication (71). Buckwheat *DFR2* differs from *DFR1* by an asparagine to valine substitution at a position that corresponds to 134 in the *Gerbera hybrida* sequence and the presence of an additional glycine between positions 138 and 139. Enzyme assays suggest that *DFR1* prefers DHK and DHM, while *DFR2* prefers DHQ. *DFR1* is expressed in most tissues, while *DFR2* expression is restricted to seeds and roots. Strawberry DFR1 has a high preference for DHK and does not accept DHQ and DHM ('DHK specialist'), while DFR2 does not accept DHK as a substrate, but works on DHQ and DHM ('DHK rejector') (72).

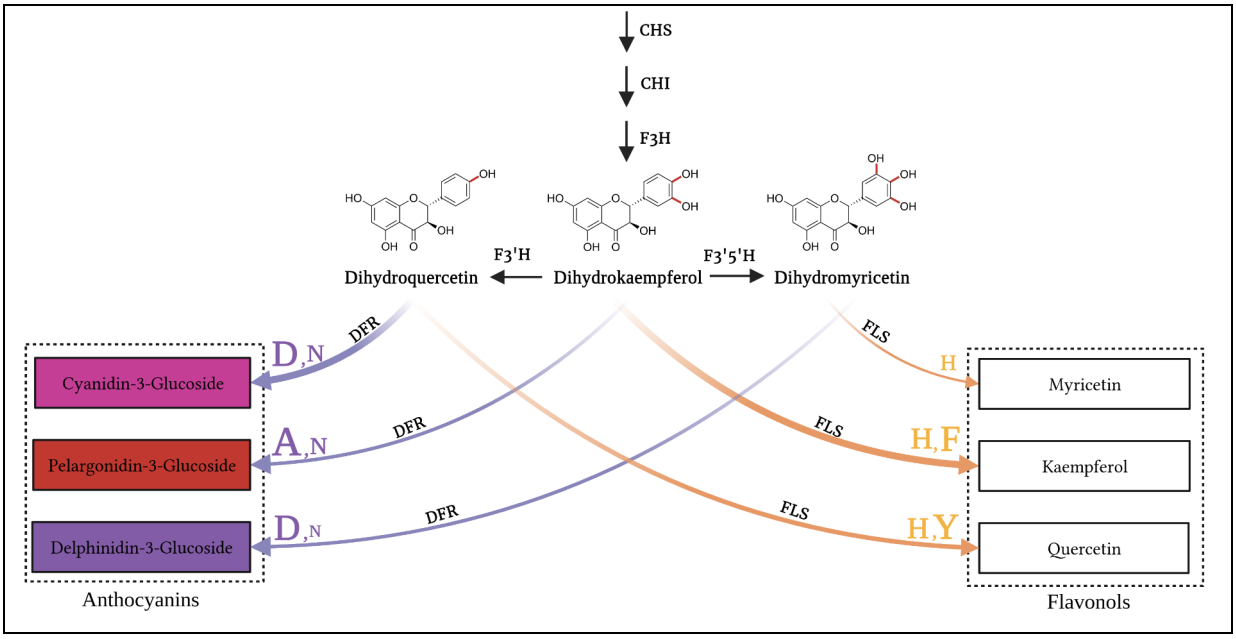

Figure 2: Illustration of the branching point in the flavonoid biosynthesis pathway leading to anthocyanins and flavonols, respectively. DFR and FLS can process substrates with different hydroxylation patterns leading to three distinct products each. The characters at the arrows represent the substrate-specificity

determining residues of FLS and DFR. The DFR residues, represented as purple characters depict the 133 position while the FLS residues, represented as yellow characters depict the 132 position in FLS (position according to *AtDFR* and *AtFLS1*).

The objective of this study was to reveal mechanisms that control the substrate competition of FLS and DFR. A systematic analysis of substrate specificity determining amino acid residues in the polypeptide sequences of FLS and DFR revealed differences between monocots and dicots. The results were connected to a cross-species gene expression analysis of relative F3'H activity which serves as a proxy for the relative availability of DHK and DHQ, respectively.

Methods

Analyzed species and data sets

The coding sequences for 211 plant species were retrieved from Phytozome (73) and NCBI/GenBank (74) (Additional File 1). The selection of plant species encompassed a diverse taxonomic range, including 20 Thallophyta (algae and fungi), 4 Bryophyta (mosses), 1 Pteridophyta (ferns and fern allies), 11 Gymnosperms (naked seed plants), and 175 Angiosperms (flowering plants). A Python script (75) was used to translate the coding sequences into corresponding peptide sequences.

Flavonoid biosynthesis gene identification

The polypeptide sequences were analyzed with KIPEs3 v0.35 (76,77) to detect the presence of flavonoid biosynthesis genes in each plant species. KIPEs3 itself supplied the bait sequences

and essential residue information required for the analysis (flavonoid biosynthesis dataset v3). By employing KIPEs3, the candidate genes for *FLS* (*flavonol synthase*) and *DFR* (*dihydroflavonol 4-reductase*) were identified. In addition, flavanone 3-hydroxylase (*F3H)*, flavonoid 3'-hydroxylase (*F3'H)*, and flavonoid 3'5'-hydroxylase (*F3'5'H)* genes were identified in the investigated plant species. KIPEs3 also enabled the identification of FLS-like and DFR-like genes within the datasets.

Multiple sequence alignment and phylogenetic tree construction

The DFR and DFR-like polypeptide sequences and the FLS and FLS-like polypeptide sequences were aligned separately using MAFFTv7.310 (78) with default penalties for gap and the protein weight matrix of BLOSUM62 (79). The amino acid alignments were translated back to a codon alignment using pxaa2cdn (80). The alignments were cleaned using pxclsq (80) to remove alignment columns with very low occupancy (>0.1). Multiple ANR and CCR sequences retrieved from GenBank were used as an outgroup for DFR tree construction. Previously characterized *DFR* genes were also included. Multiple F3H and ANS sequences obtained from GenBank were used as an outgroup for FLS tree construction to facilitate a comprehensive investigation. Previously described FLS sequences were also included. The best model of evolution was inferred using ModelFinder (81). Maximum likelihood trees were constructed for deciphering *DFR* and *FLS* evolution using IQ-TREE v2.0.3 (82) with 1,000 bootstrap replicates and the GTR+F+ASC+R10 nucleotide model. Additional phylogenetic trees were constructed based on MAFFTv7.310 and Muscle v5.1 (super5 algorithm) with IQ-TREE v2.0.3, FastTree2.1.10 (83) using the GTR+CAT model and MEGA v11 (84) using the neighbor-joining method, 1,000 bootstrap replicates and the Maximum Composite Likelihood model. The

topologies of resulting trees were manually compared to validate the crucial nodes supporting major conclusions. Phylogenetic trees were visualized using iTOLv6.8 (85).

Transcriptomic analyses

For transcriptome analyses, a subset of 43 species was selected based on specific criteria: (1) the presence of *FLS*, *DFR*, *F3H*, and *F3'H* genes in the species, (2) the availability of sufficient transcriptomic datasets in the Sequence Read Archive (SRA), and (3) the presence of clearly labeled tissue samples with a minimum of three accession numbers corresponding to at least a single tissue type (leaf, root, stem, flower, seed, or fruit). To generate the count tables, all available paired-end RNA-seq datasets of the selected species were retrieved from the SRA (www.ncbi.nlm.nih.gov/sra) (86) using fastq-dump (https://github.com/ncbi/sra-tools). Kallisto v0.44 (87) was then employed to quantify transcript abundance based on the coding sequences of the respective plant species. A previously developed Python script (88) was utilized to merge the kallisto output, resulting in one count table file for each species. The accession numbers of different tissues (leaf, stem, flower, root, seed, and fruit) for the 43 species were retrieved from the SRA. This sample selection was further reduced to avoid the inclusion of any specific stress treatments i.e. the different samples should represent different plant parts or tissues under 'normal' conditions. We aimed to assess the transcript abundance of a specific gene in a particular sample. To achieve this, a Python script (89) was employed to extract an expression data subset from the count table containing exclusively the values belonging to the accession numbers of the desired samples. The total transcript abundances (TPMs) of all isoforms of a gene for each RNA-seq sample were aggregated as previously described (90). To obtain an overall representation of the gene's transcript abundance, we calculated the median value across the row corresponding to the aggregated expression (TPM) of the gene of interest. In

cases where multiple paralogs existed within the species, the transcript abundances of these candidates for a given RNA-seq sample were combined and the median TPM value of the sums was computed to reflect the overall gene expression associated with the particular step in the biosynthesis pathway (Figure 3). The median was chosen due to its tolerance towards outliers. To visualize the gene expression patterns, a heatmap was generated using the seaborn package (91) in Python. The heatmap depicted the relative gene expression levels of *F3H*, *F3'H*, *F3'5'H*, *FLS*, and *DFR*. Additionally, ratios of *F3'H*/*F3H*, *F3'5'H*/*F3H*, and *DFR*/*FLS* were plotted to provide insights into their relative expression levels and potential interdependencies within the flavonoid biosynthesis pathway.

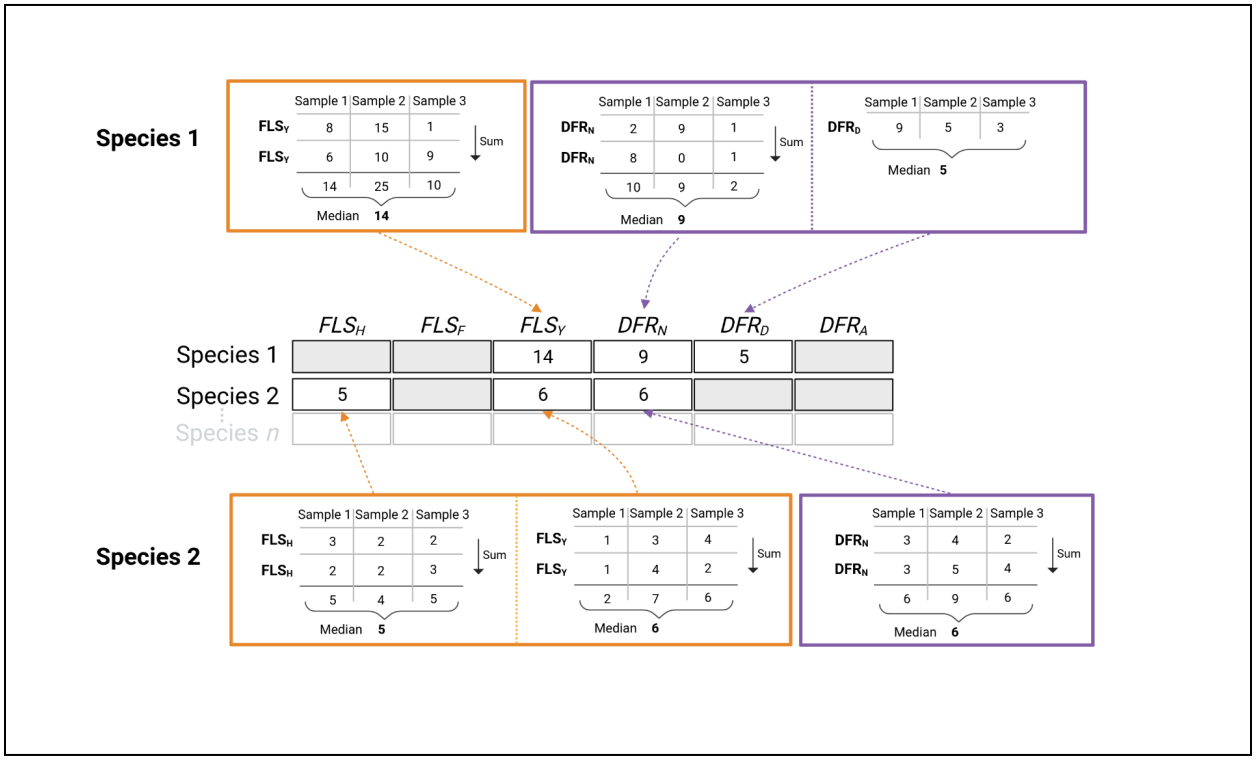

Figure 3: Illustration of the cross-species gene expression comparison. All displayed values are TPMs. The values of close paralogs were replaced by their sum if these paralogs had the same substrate affinity determining residue at the most important position.

Analyses

Substrate-specifying residues in FLS and DFR and their pattern of occurrence in land plants

From a vast range of diverse taxonomic plants, DFR was identified only in angiosperms. Among the 175 angiosperm datasets analyzed, 129 exhibited most likely functional DFR, characterized by the presence of essential residues associated with DFR activity. Overall, 207 DFR sequences were identified, accounting for cases where multiple *DFR* genes were present in certain species. Multiple sequence alignments of the identified DFR sequences revealed potential NADPH binding and substrate binding sites in N-terminal regions (**Figure 4**). Analysis based on the conservation of the amino acid residue at the 133 position i.e. 3rd position within the 26-amino acid-long substrate binding domain of DFR led to the classification of DFR proteins into three types:

(i) DFR with Asn (N) at the 3rd position which can recognize all three dihydroflavonols as substrates, (ii) DFR with Asp (D) at the 3rd position which prevents the acceptance of dihydrokaempferols, and (iii) DFR, characterized by Ala (A) at the 3rd position which exhibits a preference for dihydrokaempferol and a higher likelihood of rejecting dihydromyricetin. Henceforth, these three DFR types will be referred to as DFR_N, DFR_D, and DFR_A, respectively.

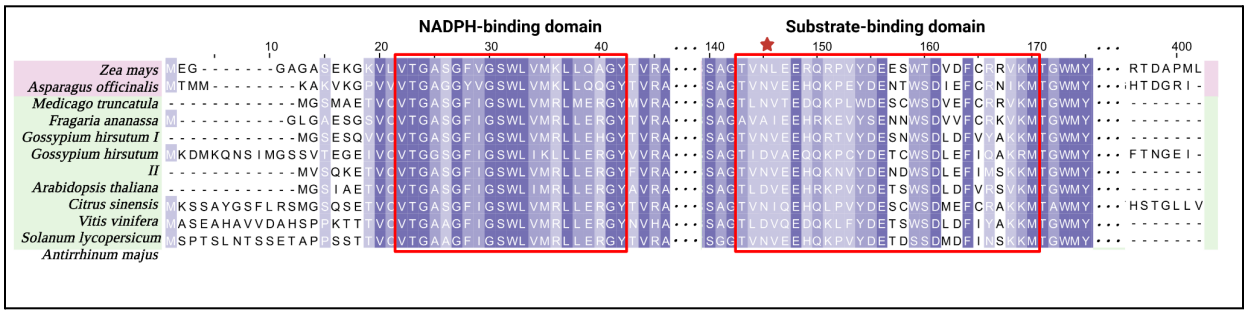

Figure 4: Selected parts of multiple sequence alignment of DFR sequences restricted to 10 species to reduce the complexity. Monocot and dicot species are highlighted by light pink and light green colour,

respectively. The dark violet, mid violet and light violet colours indicate 100%, >75%, and >50% similarity among the sequences, respectively. Important domains associated with DFR functionality are highlighted, and some columns are masked with three dots (**...**). Red boxes highlight the NADPH-binding domain and the 26AA substrate-binding domain. The red-star labeled residue at the 3rd position within this region is crucial for dihydroflavonol recognition (position 133 in the *A. thaliana* DFR). MAFFTv7 was employed to generate the alignment.

Sequences of FLS enzymes were successfully identified in the datasets of 143 out of 175 angiosperm species. The inability to identify FLS sequences in some of the plant species may be attributed to potential technical limitations of the analyzed data sets. In total, 247 FLS sequences were identified, including instances where multiple copies of FLS were present within a species. FLS enzymes belong to the 2-oxoglutarate-dependent dioxygenase (2-ODD) superfamily. A comprehensive analysis of the multiple sequence alignment revealed the presence of characteristic conserved ferrous iron-binding motif (HX(D/E)XnH), as well as 2-oxoglutarate residues (RXS) critical for FLS functionality (**Figure 5**). Furthermore, functional FLS genes exhibited additional residues (G68, H75, P207, and G261) responsible for the proper folding of the 2-ODD polypeptide. Specific motifs, such as "PxxxIRxxxEQP" and "SxxTxLVP" which were initially considered to be unique to FLS and distinguished it from other plant 2-ODDs (92) were also identified. However, a previous study identified two *BnaFLS1* homologs which also exhibited F3H activity (93). It was demonstrated that despite having these FLS-specific motifs, they could still have bifunctional activity. FLS sequences also possessed five key amino acids (H132, F134, K202, F293, and E295) identified as potential active DHQ binding site residues, with the latter four exhibiting notable conservation across diverse species. These residues were also found to be conserved in anthocyanidin synthase (ANS) sequences wherein they had a strongly conserved Y132 residue as well. Analysis based on the conservation of 132nd residue led to the classification of FLS into 3 types: FLS_H (His), FLS_F (Phe) and FLS_Y

(Tyr). The positions of amino acids or nucleotides are assigned based on the *Arabidopsis thaliana* sequences (*AtDFR*, AT5G42800 and *AtFLS1*, AT5G08640) for consistency and reference purposes.

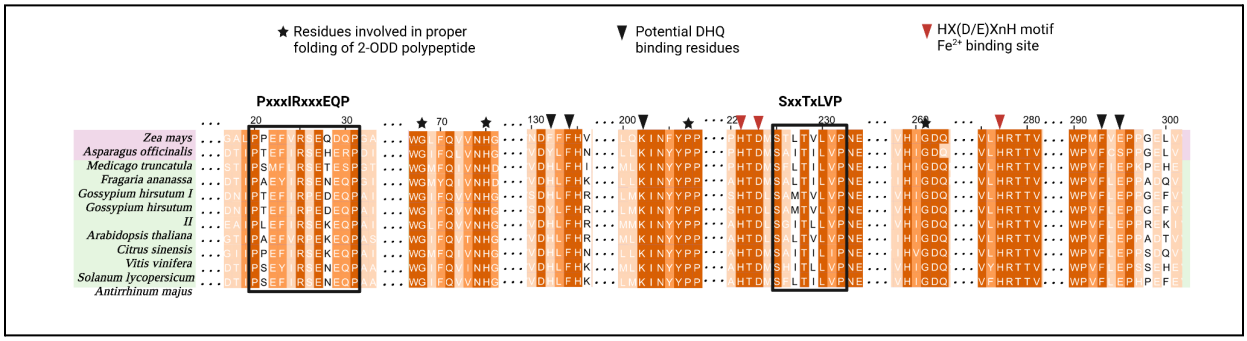

Figure 5: Selected parts of multiple sequence alignment of FLS sequences restricted to 10 species to reduce the complexity: Monocot and dicot species are highlighted by light pink and light green colour, respectively. The dark orange, mid orange and light orange colours indicate 100%, >75%, and >50% similarity among the sequences, respectively. Important domains and residues associated with FLS functionality are highlighted and some columns are omitted as indicated by three dots (**...**). The conserved 2-ODD domain and the Fe^{2+} binding sites are indicated by black asterisks and red arrowheads, respectively. The black boxes highlight the FLS-specific motifs. The black arrowheads indicate potential DHQ-binding sites where the first residue (position 132) is thought to be critical for the substrate preference of FLS. The alignment was generated by MAFFTv7.

The 129 diverse plant species harboring DFRs represented 28 orders out of the 64 orders recognized in the current Angiosperm Phylogeny Group (APG IV;2016) classification of angiosperms (94). Interestingly, all monocot plants investigated displayed DFR_N, with only a few minor exceptions which displayed neither of the three types. Among dicots, some species exclusively possessed either DFR_N or DFR_D, while others exhibited a combination of either DFR_N and DFR_D or DFR_N and DFR_A. Remarkably, out of all the plant species examined, *Fragaria* stood out as the sole genus identified to possess DFR_A, which is known to have a strong preference for DHK as a substrate, as previously reported (Additional File 2). A total of 143

angiosperm plant species encompassing 25 orders were analyzed for functional FLS genes. Among these species, monocots exhibited either FLS_F or FLS_Y, while all dicots possessed FLS_H with a histidine residue at position 132. However, certain orders such as Fabales, Fagales, Malpighiales and Malvales displayed multiple FLS candidates in some species with both FLS_H and FLS_Y, characterized by histidine and tyrosine residues, respectively. (Additional File 3). The overall pattern of amino acid residues at position 133 in DFR sequences and position 132 in FLS sequences at the order level is summarized in **Figure 6**.

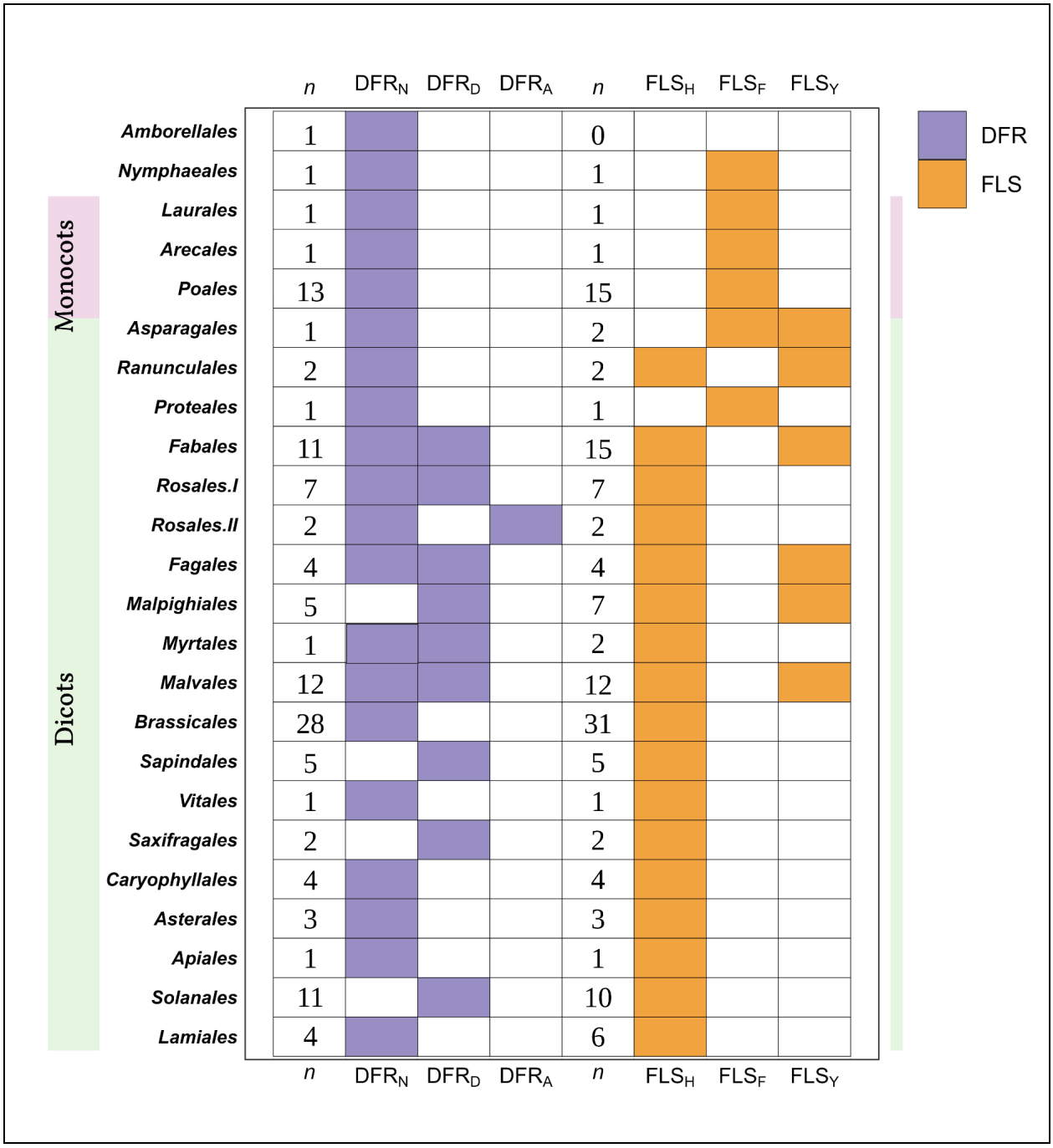

Figure 6: The patterns of substrate-determining amino acid residues observed in different orders of angiosperms for FLS and DFR. Orders are sorted by branching in the evolution of angiosperms according to the Angiosperm Phylogeny Group (2016). The residues were investigated across 172 angiosperm species and are represented here at the order level. Monocot and dicot species are highlighted by light pink and light green colour, respectively. Rosales II consists of 2 species: *Fragaria x ananassa* and

Fragaria vesca. *n* represents the number of analyzed species for DFR and FLS, respectively, within each order.

Evolution of DFR

To understand the evolutionary relationships among multiple DFR sequences, we constructed a phylogenetic tree with *DFR* protein-encoding sequences. Given that DFR belongs to the short-chain dehydrogenase (SDR) family, the closely related ANR and CCR sequences from the same family, along with DFR-like sequences were employed as outgroups. In the resulting analysis, the ANRs, and DFRs clustered together, distinct from the CCRs, as depicted in Additional File 4. The two clades preceding the CCR clade had *A. thaliana* Tetraketide α-pyrone reductase (*AtTKPR1* and *AtTKPR2*) in each clade, essential for pollen wall development and male fertility. TKPR2, known to be more active than TKPR1, clustered closer to the CCR clade. The TKPR1-like, TKPR2-like, CCRs and a large clade with an unknown function clustered together and was a sister clade to the DFR and ANR clade. Adjacent to the DFR clade, another clade featured AtBEN1-encoding a DFR-like protein regulating brassinosteroids in *A. thaliana*. This clade did not contain any sequences from the gymnosperms or monocots, only dicots were found. Intriguingly, this clade, exclusive to dicots, exhibited a diverse composition of species besides brassicales, challenging its presumed species-specific function. The proximity to the DFR clade and the absence of monocots suggests its evolutionary origin from a DFR predecessor gene.

We did not identify any Gymnosperm DFR in our analysis, possibly due to their early-branching nature and distinct conserved residues compared to angiosperms. Previously characterized DFR sequences (Additional File 5), marked with red asterisks, were generally located within the DFR clade, except for fern *Dryopteris erythrosora*, *DeDFR1*, and *DeDFR2*, *Ginkgo biloba*, *GbDFR1*, *GbDFR2*, and *GbDFR3*, and *Camellia sinensis*, *CsDFRc*, which were outside, generally indicating non-DFR sequences. Fern DFRs, found before the ANR and DFR clade,

may exhibit distinct characteristics due to their early-branching nature or possess promiscuous functions. Ginkgo DFRs within TKPR-like clades likely deviate from true DFRs. The phylogenetic analysis of functional DFRs highlights distinct clades specific to monocots and dicots (**Figure 7**). Monocots exclusively exhibit DFR_N, while dicots generally display various DFR types - DFR_N, DFR_D, and DFR_A. Notably, DFR_D appears to have originated from a duplication event of DFR_N, as observed in the Fabales and Malvales clades. Within the Rosales clade, a unique residue type, Alanine, is present, likely arising from a substitution event of the N residue. These findings suggest a possible scenario where dicots experienced a duplication event in DFR_N, followed by a subsequent substitution event resulting in the emergence of DFR_D. Furthermore, in the Rosales clade, the duplication was accompanied by a substitution from N to A.

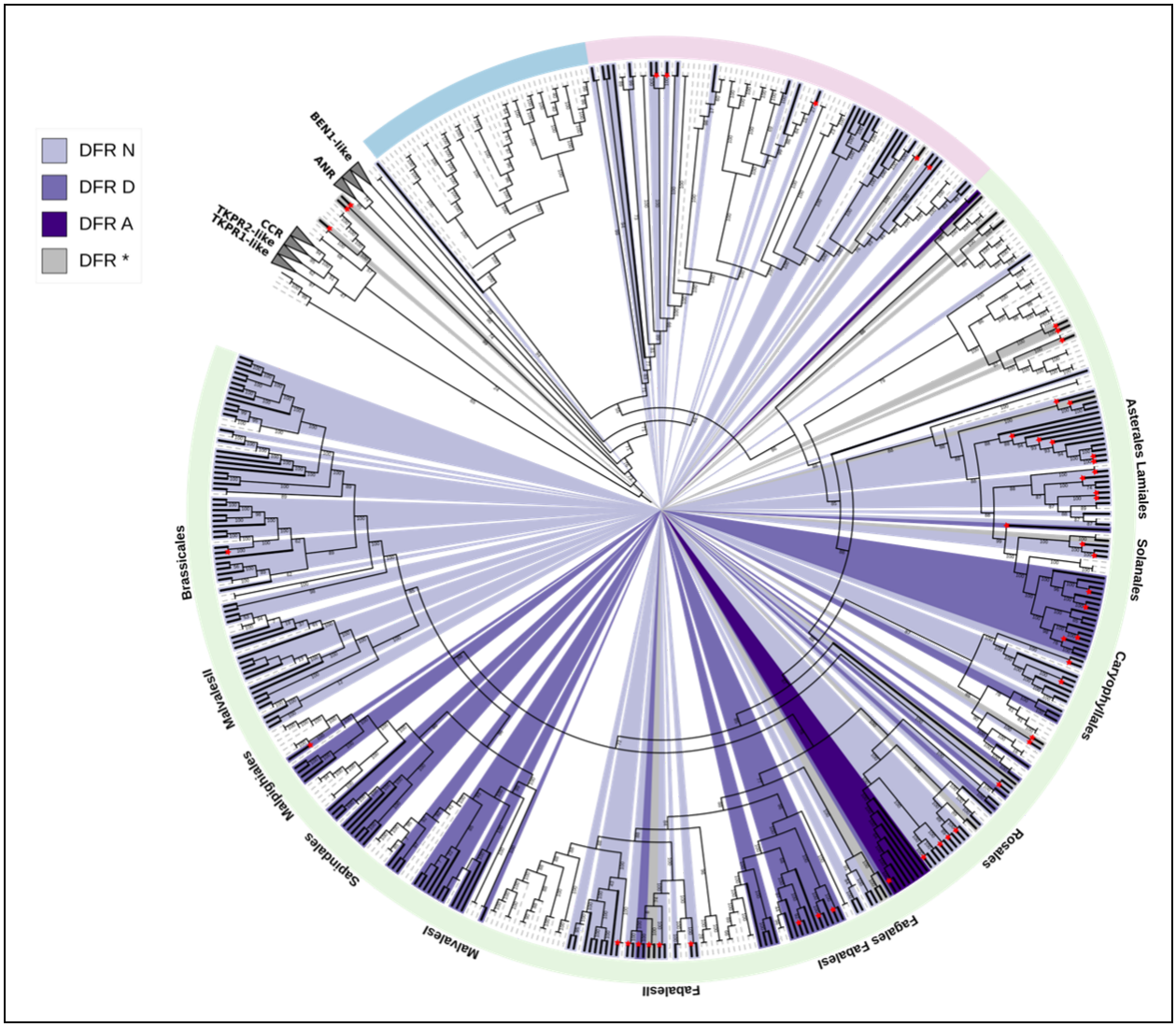

Figure 7: Phylogenetic analysis of DFR sequences in diverse plant species, highlighting amino acid residue diversity at position 133 associated with substrate specificity. Gymnosperm, monocot and dicot species are denoted by light blue, light pink and light green colour stripes, respectively. Non-DFR sequences are represented by dashed gray branches while the functional DFRs are indicated by solid black branches. The color-coded scheme represents different residues: Asparagine (light purple), Aspartic acid (periwinkle blue), Alanine (deep purple), and other amino acids (grey) at position 133. Distinct clusters of DFRs from major plant orders are labeled for reference. DFR genes characterized in previous studies are highlighted by a red asterisk at the start of the node branch. Leaf labels are hidden to reduce the complexity. The outgroup comprises SDR members like ANRs, CCRs and other DFR-like sequences (Additional File 4).

Deep duplication events during the evolution of FLS

Through a phylogenetic tree generated with *FLS* protein-encoding sequences, we observed deep duplication events during the evolution of FLS enzymes. Considering their classification within the 2-oxoglutarate-dependent dioxygenase (2-ODD) family, closely related ANS, F3H and FNS1 sequences from the same family and FLS-like sequences lacking crucial residues for FLS functioning were included as outgroups. The phylogenetic tree revealed a largely distinct clustering of F3H, ANS and FLS sequences. ANS and FLS formed a common cluster, while F3H grouped with FNS1 sequences (Additional File 6).

The non-Apiaceae FNS1 sequences clustered outside the F3H clade while the liverworts FNS1 and Apiaceae FNS1 were a part of a single clade with phylogenetically distinct F3H sequences. The liverworts FNS1 are present at the root of this clade suggesting that plant F3H evolved from liverworts FNS1. An independent lineage of Apiaceae FNS1 is located nested within the Apiaceae F3H sequences further confirming the evolution of Apiaceae FNS1 from F3H (88,95). Gymnosperm, Monocot and dicot sequences clustered separately in the four 2-ODD members, i.e. F3H, ANS and FLS. The FLS clade was divided into 4 primary clusters: a Gymnosperm clade branching first, a second clade which we hereafter call aFLS (ancestral FLS), a third monocot clade and lastly, a large dicot clade. The aFLS clade encompasses FLS-like sequences from both monocots and dicots and a few dicot FLS sequences. Only 9 FLS sequences, one each from *Linum usitatissimum, Lotus japonicus, Cajanus cajan, Quercus suber, Betula platyphylla, Acacia confusa AcFLS*(JN812062), *Camellia sinensis CsFLSa*(KY615705), *Fagopyrum tataricum FtFLS2*() *and Vitis vinifera FLS5*(AB213566) were in this clade, although the FLS-like sequences were from a wide range of plants. The presence of this aFLS clade sister to the distinct monocot and dicot clade implies an FLS duplication event preceding the monocot-dicot split. The large dicot cluster further separates into two distinct subclades: one including sequences only from the Brassicales order, and the other including sequences from the rest of the dicot plants (**Figure 8**). In the Brassicales subclade, a

monophyletic group of Brassicales FLS and FLS-like sequences is evident, with *AtFLS1* and other functional Brassicales FLSs clustered together. The presumed non-functional FLS sequences from the Brassiclaes order group together and are classified based on their phylogenetic relationship to their most likely *A. thaliana* orthologs (*AtFLS2*-*AtFLS6*). Within the latter subclade which excludes Brassicales, some FLS sequences from the same order appear in distinct clusters as observed in the Fagales, Sapindales, Malpighiales and Malvales order, suggesting gene duplication events in the evolutionary history of FLS in some common ancestor of these orders.

While the first and last residues of the FLS substrate binding residues (H132, F134, K202, F293, and E295) exhibit diverse residues, the remaining three are strongly conserved in both FLS and ANS sequences. Notably, gymnosperm FLSs and aFLSs feature a Y132 residue, consistent with ANS sequences. Monocotyledonous plants display an F/Y(85%/15%) residue at position 132, whereas the majority of dicots consistently possess an H residue at position 132, except for multiple *Gossypium* species and *Herrania umbratica* from the Malvales order, which have a Tyrosine residue at this position. The E295 residue is conserved in both ANS and monocot/dicot FLS. However, gymnosperms, aFLSs, AtFLS4, and AtFLS5 clades exhibit non-conserved amino acid residues at position 295. The aFLSs exclusively feature aliphatic amino acids at this position (G/A/V/L/I).

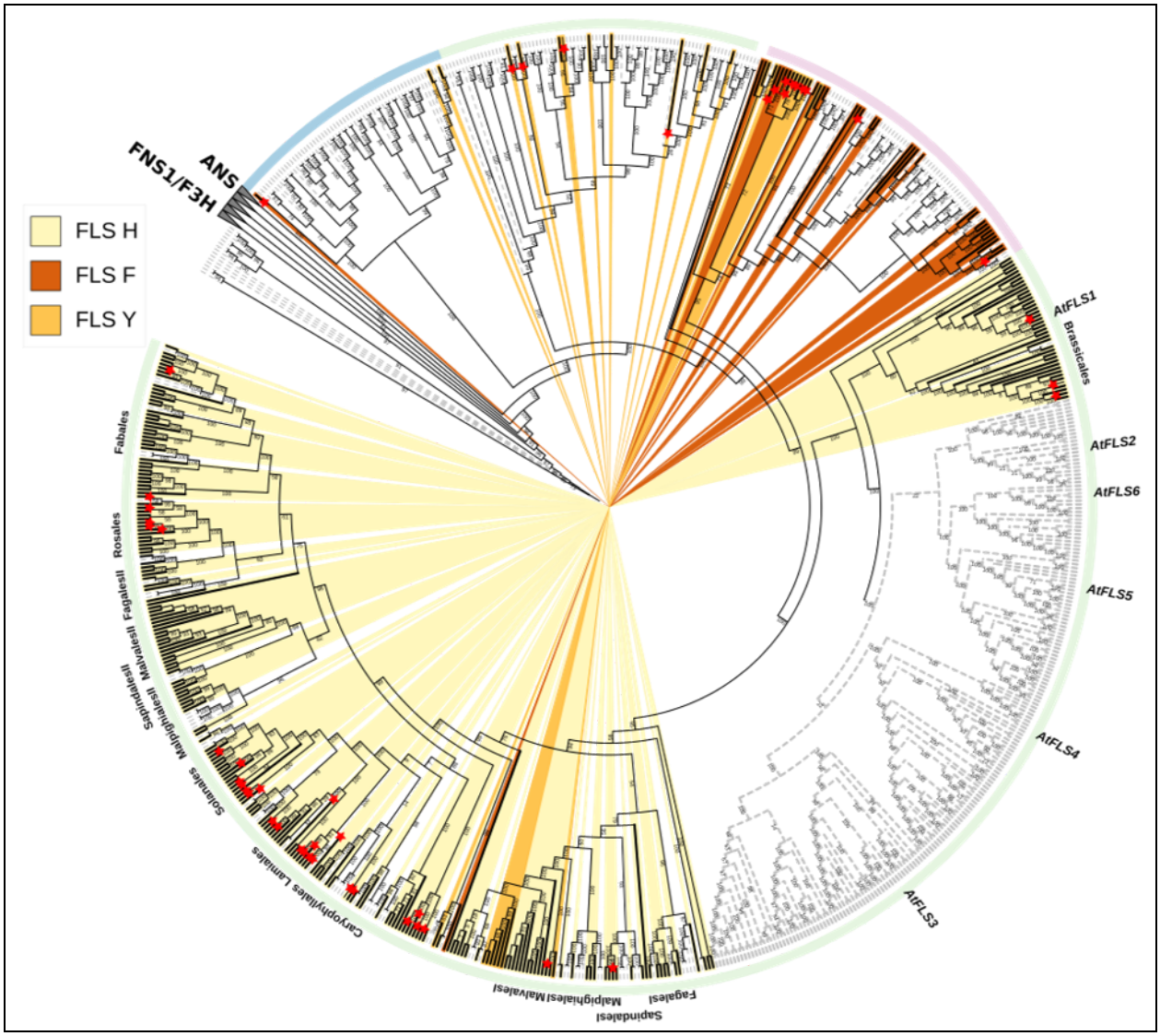

Figure 8: Phylogenetic diversity of FLS sequences in diverse plant species, highlighting amino acid residue diversity at position 132 associated with DHQ-substrate binding. Gymnosperm, monocot and dicot species are denoted by light blue, light pink and light green colour strips, respectively. Non-FLS sequences are represented by dashed gray branches while the functional FLSs are indicated by solid black branches. The background colour highlights different presumably substrate-determining amino acid residues: histidine (pale yellow), phenylalanine (dark orange), and tyrosine (dark golden) at position 132. Distinct clusters of FLSs from major plant orders are labeled for reference. FLS genes characterized in previous studies are highlighted by a red asterisk at the start of the node branch. The branches of

Arabidopsis thaliana AtFLS1-AtFLS6 are labeled. Leaf labels are hidden to reduce the complexity. The outgroup comprises members of 2-ODD like F3H, ANS and other FLS-like sequences (Additional File 6).

Inferring dihydroflavonol availability from relative *F3H*, *F3'H*, and *F3'5'H* gene expression

Besides *FLS* and *DFR*, the major genes influencing the branching into flavonol and anthocyanidin biosynthesis are *F3H*, *F3'H*, and *F3'5'H*. These genes play a crucial role in determining the hydroxylation patterns of dihydroflavonols, which are common substrates for FLS and DFR. F3H catalyzes the formation of DHK, while F3'H adds a hydroxy group converting DHK into DHQ. F3'5'H catalyzes the formation of triple hydroxylated DHM from DHK/DHQ. Understanding the substrate preference of FLS and DFR for differently hydroxylated substrates requires information about the relative availability of DHK, DHQ, and DHM. Since metabolic datasets are not available at the necessary scale, we utilized gene expression data as a semi-quantitative proxy for metabolomics. The relative gene expression of *F3H*, *F3'H*, and *F3'5'H* is analyzed to infer the relative production of the three flavonol/anthocyanin precursors i.e. DHK, DHQ and DHM. The ratio of *F3'H* to *F3H* and the ratio of *F3'5'H* to *F3H* were calculated across 43 plant species. For example, a high *F3'H* to *F3H* ratio would suggest that DHQ accounts for most of the dihydroflavonols. These gene expression ratios were connected to the expression levels and substrate-determining amino acid residues of *FLS* and *DFR*.

To investigate the gene expression patterns, a heatmap was generated to depict the transcript abundance of the genes in 43 species representing the taxonomic diversity of angiosperms (Additional File 7). Here, we present 11 species to reduce the complexity. The heatmap covers several tissue samples such as leaf, root, stem, flower, fruit, and seed. The samples displayed per individual species depend on data availability. The heatmap analysis revealed distinct expression patterns of *FLS* and *DFR* genes in different plants. In Figure 9, the monocot, *Miscanthus sinensis*, demonstrates a higher *F3'H* to *F3H* and *DFR* to *FLS* ratio, suggesting a

potential predominance of dihydroquercetin utilization by DFR for anthocyanin or proanthocyanidin production. In *Vigna unguiculata*, red fruit exhibits significant upregulation in *F3H*, *F3'H*, DFR_N, and DFR_D compared to green fruit, while *FLS* expression remains relatively constant. DFR_D prefers DHQ while the elevated *F3H* level indicates increased DHK, likely utilized by DFR_N, leading to the production of cyanidin and pelargonidin, contributing to the observed red phenotype. In *Fragaria x ananassa*, heightened *F3H* and *F3'H* expression and the absence of *F3'5'H*, suggest abundant DHK and DHQ substrates. The presence of DFR_A (Ala) and high *F3H* transcript abundance in fruit, and high *FLS* transcript abundance in flowers aligns with the observed red fruit and white flower characteristics. Cotton exhibits multiple FLS and DFR candidates of FLS_H, FLS_Y, DFR_N, and DFRD type. Its FLS_Y, limited to the Malvales order in the dicot clade, has low transcript abundance. In cotton, high expression of *F3H*, *F3'H*, *F3'5'H*, FLS_H, DFR_N, and DFR_D signifies active flavonoid pathway engagement. *Arabidopsis thaliana* lacks *F3'5'H* and exhibits relatively low *DFR* transcript abundance. The *F3H* and *FLS* transcript abundance follow the same patterns, suggesting a preference for DHK as a substrate. Species with exclusive DFR_D (Asp), like oranges and tomatoes, exhibit low *DFR* transcript abundance despite significant *F3'H* expression.

Gene Expression (TPMs)
0
4
1030
Ratio of Gene Expression (TPMs)
0
1
230
F3H
F3'H
F3'5'H
FLSH
FLSF
FLSY
DFRN
DFRD
DFRA
F3'H/F3H
F3'5'H/F3H
DFR/FLS
Miscanthus sinensis
1.42 41.01 17.85 7.01 0.89 28.92 12.59 0.13 Leaf (n=26)
0.29 48.46 15.26 5.43 0.42 166.69 52.50 0.08 Stem (n=7)
Lotus japonicus
226.56 14.42 1.99 63.84 0.01 10.35 1.89 0.06 0.01 0.78 Leaf (n=3)
547.25 518.88 22.12 152.76 0.05 43.74 47.29 0.95 0.04 0.61 Flower (n=21))
43.52 35.23 3.47 10.94 0.33 19.03 2.95 0.81 0.08 6.34 Root (n=65)
1021.40 60.71 6.68 389.18 0.02 11.78 10.90 0.06 0.01 0.11 Stem (n=6)
40.55 42.26 0.08 82.98 0.55 13.05 0.24 1.04 0.00 0.72 Seed (n=23)
629.41 54.53 106.90 74.45 0.01 178.25 110.55 0.09 0.17 3.90 Fruit (n=3)
Vigna unguiculata
38.65 28.33 0.04 8.06 0.01 0.53 0.73 0.00 0.08 Leaf (n=39)
67.77 24.13 74.34 31.25 3.85 5.21 0.36 1.10 0.29 Flower (n=3)
4.08 1.38 2.13 0.01 0.92 2.56 0.34 0.52 344.64 Root (n=14)
47.45 9.07 2.74 10.19 1.41 2.32 0.19 0.06 0.36 Stem (n=3)
11.42 13.80 2.66 3.33 1.35 2.03 1.21 0.23 1.02 Seed (n=12)
1.89 14.86 0.13 9.44 0.07 2.03 7.87 0.07 0.07 Fruit_Green (n=3)
33.84 21.16 0.33 9.40 26.40 11.24 0.63 0.01 4.10 Fruit-Red (n=6)
Glycine max
30.91 15.99 0.01 35.83 0.20 0.52 0.00 0.01 Leaf (n=158)
524.66 18.35 0.09 172.14 29.37 0.03 0.00 0.17 Flower (n=10)
17.97 17.61 0.01 0.05 0.01 0.98 0.00 0.20 Root (n=41)
115.73 37.37 0.01 27.10 33.91 0.32 0.00 1.25 Stem (n=15)
5.67 7.46 0.01 2.56 0.21 1.32 0.00 0.08 Seed (n=109)
116.07 44.69 0.01 22.84 1.67 0.39 0.00 0.07 Fruit (n=4)
Fragaria x ananassa
119.20 87.25 12.19 66.32 3.75 0.73 5.85 Leaf (n=13)
99.71 56.97 134.94 39.97 6.80 0.57 0.36 Flower (n=15)
342.46 11.76 0.90 135.43 183.03 0.03 348.95 Fruit (n=80)
Gossypium hirsutum
647.78 131.06 213.83 271.36 0.31 158.77 232.09 0.20 0.33 1.54 Leaf (n=36)
489.98 74.48 107.03 42.59 0.02 271.46 114.82 0.15 0.22 9.37 Flower (n=27)
124.72 26.12 71.45 0.30 0.01 37.07 59.68 0.21 0.57 319.21 Root (n=7)
91.54 18.45 44.71 2.97 0.05 44.39 34.05 0.20 0.49 22.86 Stem (n=5)
64.35 12.89 10.95 2.38 0.01 6.87 13.76 0.20 0.17 8.66 Seed (n=7)
758.91 12.50 152.41 4.52 0.04 186.21 342.75 0.02 0.20 102.22 Fruit (n=11)
Arabidopsis thaliana
4.50 0.25 6.77 0.31 0.05 0.05 Leaf (n=71)
74.89 20.84 155.54 0.08 0.28 0.00 Flower (n=16)
51.41 3.08 21.89 0.01 0.06 0.00 Root (n=23)
18.07 1.85 3.88 7.19 0.10 1.85 Stem (n=14)
4.50 0.25 6.77 0.31 0.05 0.05 Seed (n=7)
Citrus sinensis
2.18 8.05 0.01 85.83 0.01 3.70 0.00 0.00 Leaf (n=21)
0.71 40.73 0.01 193.74 0.01 57.75 0.01 0.00 Root (n=7)
34.52 32.06 0.01 220.05 0.01 0.93 0.00 0.00 Fruit (n=24)
Vitis vinifera
468.57 177.34 98.55 25.04 110.46 0.38 0.21 4.41 Leaf (n=11)
232.54 134.39 4.79 59.83 67.71 0.58 0.02 1.13 Flower (n=4)
60.51 48.81 9.74 0.12 35.10 0.81 0.15 293.01 Stem (n=8)
557.48 186.05 2.61 0.39 170.69 0.09 0.02 432.76 Seed (n=3)
222.85 134.39 1.48 7.81 67.18 0.60 0.01 8.61 Fruit (n=7)
Beta vulgaris
15.22 1.76 4.73 0.01 0.12 0.00 Leaf (n=16)
26.95 8.86 0.05 0.01 0.33 0.20 Root (n=36)
Solanum lycopersicum
3.26 9.19 0.28 12.89 0.43 2.82 0.08 0.03 Leaf (n=22)
154.47 11.85 0.41 127.70 0.35 0.08 0.00 0.00 Flower (n=4)
1.36 15.41 0.01 1.48 0.05 11.34 0.01 0.03 Root (n=10)
10.74 52.88 0.19 49.79 0.66 4.93 0.02 0.01 Seed (n=4)
48.07 5.07 0.02 112.34 0.19 0.11 0.00 0.00 Fruit (n=14)
F3H
F3'H
F3'5'H
FLSH
FLSF
FLSY
DFRN
DFRD
DFRA
F3'H/F3H
F3'5'H/F3H
DFR/FLS
0
4
1030
Gene Expression (TPMs)
0
1
230
Ratio of Gene Expression (TPMs)

Figure 9: Transcript abundance heatmap and gene ratio analysis of 11 selected species across various tissue samples. White tiles indicate that no gene was discovered for a certain function/type. The heatmap tiles represent the actual expression values in TPMs, with the number of samples indicated on the right. The expression heatmap shows genes associated with the branching of *FLS* and *DFR* within the flavonoid biosynthesis pathway. Blue and yellow colours indicate high and low expression levels, respectively (see colour bar). On the right, a heatmap illustrates the ratios of gene expression to investigate the substrate hydroxylation pattern in the FLS-DFR branching. Dark and light colours indicate high and low expression levels, respectively, based on the colour bar. A full heatmap with all 43 investigated species is available in Additional File 7.

Discussion

Factors influencing the competition of the pivotal enzymes FLS and DFR for the common substrate, dihydroflavonols, remain poorly understood. The equilibrium between these enzymes can be modulated by (i) the unique substrate specificity exhibited by FLS and DFR for different dihydroflavonols, (ii) the existence of multiple FLS/DFR copies with diverse substrate preferences, or (iii) the expression of specific FLS/DFR copies in specific tissues. The availability of different dihydroflavonols is determined by the activity of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase.

Mitigating substrate competition between FLS and DFR

A well-studied 26-amino acid region determines substrate specificity in DFR (50,96) and the grape crystal structure further emphasized the significance of the N133 amino acid in binding the 3' and 4' hydroxyl groups of DHQ (58). While the substrate specificity of DFR has been extensively studied, the substrate specificity determining residues of FLS have received

comparatively less attention. The substrate-binding residues of FLS (H132, F134, K202, F293, E295) show high conservation among all plant FLSs, with the exception of the H132 and E295 residues. Through the examination of *A. thaliana fls1* mutants, it was hypothesized that the H132 residue may play a critical role in governing the diverse substrate preferences of FLS enzymes (68). Numerous previous studies have employed enzyme assays to investigate the substrate preferences of plant FLSs and DFRs, while many others identified the type of anthocyanins and flavonols in various plants. Notably, *AcFLS* from the monocot *Allium cepa*, characterized by the Y132 residue, exhibits a preference for DHQ over DHK (64), while monocots such as *OsFLS* (59) and *ZmFLS* (60), with the F132 residue, favor DHK over DHQ. In contrast, DFR_N in maize and *Dendrobium officinale* predominantly produces cyanidin derivatives (97,98). DFR_N of barley displays *in vitro* reduction of DHK and DHQ whereas DHM was not tested (99), however in the testa tissue of grain procyanidin and prodelphinidin were present and leucopelargonidin and its derivatives were not found (100). DFR_N of grape hyacinth and *Cymbidium* hybrids exhibit high activity towards DHM and DHQ (96,101), and *Anthurium andraeanum* DFR, featuring an uncommon S133 residue, displays a marked specificity for DHK (102).

Among dicots, multiple DFR copies exist; for instance, in the Fabales order, *M. truncatula* DFR1 and *L. japonicu*s DFR2 and DFR3 with an N133 residue prefer DHK over DHQ, while *M. truncatula* DFR2 and *L. japonicus* DFR5 with a D133 residue show a greater reduction in DHQ (56,57). In the Rosales order, FLS_H from *Fragaria ananassa* (103), *Malus domestica* (104), and *Rosa hybrida* (105) reduces both DHK and DHQ, with strawberry FLS favoring DHK and rose FLS favoring DHQ. Conversely, their DFR_N exhibits a higher affinity for DHQ than DHK (72,106,107). The strawberry species, *Fragaria ananassa* and *Fragaria vesca* have another DFR-type, DFR_A which has a very high specificity for DHK (72). Other examples are *L. japonicus* and *M. truncatula*, where a DFR_D is present with a high preference for DHQ, while the DFR_N prefers DHK. These observations suggest a neofunctionalization of novel DFR copies i.e. the

evolution of a different substrate specificity. Exploring the spatiotemporal expression of novel *DFR* copies might help to resolve their specific functions.

In the Brassicales, FLS_H from *A. thaliana* and *Matthiola incana* prefer DHK (61,108,109), while their DFR_N reduces both DHK and DHQ but favors DHQ (110–112). *Citrus unshiu* FLS_H has higher DHK activity (62), whereas *Zanthoxylum bungeanum* from the same order, Sapindales, has DFR_D and prefers DHM followed by DHQ and DHK (113). In caryophyllales, FLS_H in *Dianthus caryophyllus*, *Fagopyrum dibotrys*, and *Fagopyrum tataricum* catalyze both DHK and DHQ (65,114,115), while DFR_N in *Dianthus* reduces DHQ more effectively than DHK (114). In *Fagopyrum esculentum*, DFR_N, active in all tissues, reduces both DHK and DHM at the same rate, while the second DFR enzyme with a V134 residue, expressed only in roots and seeds, exhibits high specificity towards DHQ (71). In Solanales, *Petunia hybrida* and *Nicotiana tabacum* FLS_H catalyze both DHK and DHQ with minimal activity towards DHM (53,116), whereas their DFR_D is specific in reducing 3,4 di- and 3,4,5 tri-hydroxylated substrates but cannot reduce DHK (50,51). *Gentiana triflora* predominantly produces blue flowers and generally have cyanidin, delphinidin, and quercetin derivatives (117).

In our study, we observed that all monocots were mostly FLS_F type, and some were FLS_Y type, while their DFR were predominantly DFR_N type. These enzyme assay studies suggest that in monocots, FLS_F generally prefers DHK while DFR_N prefers double or triple hydroxylated dihydroflavonol substrates. This difference in substrate preferences could help in avoiding competition between the two enzymes. In dicots, FLS_H is observed in the majority of plants, but in fewer plants with multiple FLS sequences, FLS_Y is also observed. While DFR is more diverse in dicots, some have only DFR_N or only DFR_D, while others have DFR_N along with either DFR_D or DFR_A candidates. The prior studies collectively suggest that FLS_H in dicots can utilize both DHK and DHQ as substrate but may have a preference for one over the other in different plants, implying that the H132 residue remains more neutral between DHK and DHQ. The biosynthesis of DHM requires the presence of flavonoid-3',5'-hydroxylase (F3'5'H) activity, which is rare.

Since blue flower color can attract bee pollinators, it is only found in certain angiosperm lineages (118). DHQ and DHM (if F3'5'H gene is present) appear to be the more preferred substrate of DFR in most cases. Both in monocots and dicots, double and triple hydroxylated substrates seems to be more preferred by DFR, aligning with the notion that cyanidin is the most abundant and widely distributed floral anthocyanin (119). While substrate-specifying residues may play a role in identifying the specificity of FLS and DFR enzymes, it is plausible that substrate preference may not be exclusively determined by the H132 and N133 residues, respectively, in all plants.

Functional divergence during FLS evolution

In the plant kingdom, the 2-oxoglutarate-dependent dioxygenase (2-ODD) superfamily is one of the largest enzyme families only second to cytochrome P450s (CYPs). Within this superfamily, the plant 2-ODD family can be categorized into three classes: DOXA, DOXB, and DOXC, with all enzymes involved in flavonoid biosynthesis falling into DOXC. These 2-ODD enzymes are believed to have originated from a common ancestor before the emergence of land plants, and then underwent species-specific evolution in response to diverse environmental conditions (124). The evolution of FLS and ANS is likely to have occurred after the emergence of F3H during seed plant evolution (124). A noteworthy observation similar to that reported by Wang et al. (2019) (125), is the dispersion of ginkgo 2-ODD genes (*GbFLS*, *GbANS*, and *GnF3H*) into FLS, ANS, and F3H groups, respectively, indicating their divergence preceding the separation of gymnosperms and angiosperms.

The large dicot clade comprises a Brassicales order subclade and another clade comprising the rest of the dicots. Both clades show multiple duplication patterns at a shallower level. We presume that sequences in the Brassicales clade have functions similar to their *A. thaliana*

AtFLS1-AtFLS6 orthologs. Only *AtFLS1* is considered a functional *FLS* gene. *AtFLS4* and *AtFLS6* are non-functional pseudogenes and *AtFLS2*, *AtFLS3*, and *AtFLS5* have been reported to retain partial FLS activity under specific conditions (126). However, other studies only found strong evidence for FLS activity in the Brassicales FLS1 clade (61,93)). Given that the FLS2-FLS6 clade is present across multiple Brassicales species, it is likely that they have a yet undiscovered function. Distinct E295 residue patterns are observed among the FLS clades; AtFLS1, AtFLS2, AtFLS3, and AtFLS6 have E/D295 residue, whereas AtFLS4 and AtFLS5 have other non-conserved residues at this position. The AtFLS5 clade features aliphatic amino acids at this position, and *AtFLS5* has been reported to be expressed only in the roots of seedlings (61). Enzyme assay results by Chua et al. (2018) indicate that mutations H132Y and H132F show higher specific activity than wild-type Arabidopsis FLS1, with E295L mutation resulting in only 7% of WT activity. However, the study only used DHQ as a substrate, and DHK was not tested. Additional studies testing enzyme activity of mutant FLS (H132Y, H132F, and E295(V/L)) with different substrates would help confirm the function of these aFLSs and the substrate preferences of FLSs with different residues.

In the second subclade, an interesting evolutionary pattern is observed, where FLS sequences from some plant orders appear in two different clades, particularly noticeable in the Fagales, Malpighiales, Malvales, and Sapindales orders. This suggests significant divergence in FLS sequences in these orders due to functional requirements, leading to the formation of distinct clades. Within the Malvales order, including cotton species, an additional duplication event was observed, originating from FLS_H duplication, followed by a subsequent substitution leading to the re-emergence of FLS_Y (H132Y). The exclusive presence of the Y residue at position 132 in the Malvales order indicates a potential evolutionary step to enhance substrate affinity. Other orders like Solanales, Rosales, and Fagales are only represented in a single FLS clade.

Although the *Citrus unshiu* FLS within the Sapindales I clade exhibits a significantly greater affinity for DHK compared to DHQ, the satsuma mandarin predominantly accumulates quercetin

3-O-rutinoside (rutin) (127). The authors of this study proposed the presence of more than one FLS in *C. unshiu* (62). In other *Citrus* species, like *C. sinensis*, *C. trifoliata* and *C. clementin*a, we observed two FLS candidates, one each in the Sapindales I and II clade. Our analysis and enzyme assay data from previous studies suggest that FLS sequences within the first dicot subclade (Fagales I, Sapindales I, Malpighiales I, and Malvales I) display a pronounced substrate specificity for DHK. In contrast, FLS sequences within the second dicot subclade show a preference for both DHK and DHQ and may have a higher specificity for DHQ (63,65,66,103,105,114,128).

ANS/LDOX, which catalyses the immediate step downstream of DFR, is closely related to FLS and shows 50-60% protein sequence similarity. Besides the conserved 2-ODD family residues, ANS/LDOX and FLS have similar DHQ binding residues (Y132, F134, K202, F293, and E295 conserved ANS/LDOX residues based on position in AtFLS1) and the substrate/ascorbate binding residue of ANS/LDOX, E230 in AtLDOX. Bifunctional ANS enzymes with FLS-like activity have been discovered in *G. biloba* (120), *O. sativa* (121), *A. thaliana* (69), and *M. truncatula* (122). *In vitro* studies have shown that ANS can convert leucocyanidin to cyanidin and also dihydroquercetin to quercetin. *At*LDOX (123) converted DHQ and DHK into quercetin and kaempferol, respectively, in the ratio of 1:0.7 suggesting a higher affinity for DHQ at least in *A. thaliana*. This suggests that if FLS is specific for DHK, ANS might be responsible for different patterns of accumulation of kaempferol and quercetin observed in different plants. While DFR is more likely to accept DHQ, the ANS also might be more adapted to produce leucocyanidins.

Conclusion

The branching point of the flavonoid biosynthesis into flavonols (FLS) and anthocyanins/proanthocyanidins (DFR) is a pivotal step for metabolic flux. The availability of

massive plant sequence datasets and gene expression resources enabled the investigation of functional FLS and DFR sequences across a wide spectrum of plant species. This transformative approach revealed global patterns that cannot be identified by analyzing individual gene functions in one species through classical wet lab methods. We were able to pinpoint putative substrate-determining residues, revealed evolutionary patterns, and shed light on potential competition mitigation mechanisms. These insights into competing branches responsible for producing coloured anthocyanins and non-coloured flavonols can guide molecular breeding or engineering of ornamental plants.

Declarations

Data Availability

Scripts developed for this project and additional datasets are available on GitHub (https://github.com/bpucker/DFR_vs_FLS). The count tables of all 43 species analyzed via RNA-seq are freely available through LeoPARD (https://leopard.tu-braunschweig.de/receive/dbbs_mods_00073263) (129).

Funding

NC received a DAAD-Kospie scholarship for her research stay at TU Braunschweig which enabled her to conduct most of the presented analyses. NC's current position is funded by the DFG - Project number 516566273.

Competing interest

The authors declare that they do not have any competing interests.

Authors' contribution

NC and BP designed the experiments, collected data, wrote scripts for the data analysis, generated figures, and wrote the manuscript. All authors read and approved the final version of this manuscript.

Acknowledgements

We thank the research group Plant Biotechnology and Bioinformatics at TU Braunschweig for excellent discussions. This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A). We acknowledge support by the Open Access Publication Funds of Technische Universität Braunschweig.

References

1. Winkel-Shirley B. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiol. 2001 Jun 1;126(2):485–93.
2. Koes R, Verweij W, Quattrocchio F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. 2005 May 1;10(5):236–42.
3. Paz-Ares J, Ghosal D, Saedler H. Molecular analysis of the C1-I allele from Zea mays: a dominant mutant of the regulatory C1 locus. EMBO J. 1990 Feb;9(2):315–21.
4. Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E. Control of anthocyanin biosynthesis in flowers of Antirrhinum majus. Plant J. 1991 Jul;1(1):37–49.
5. Buer CS, Muday GK. The transparent testa4 Mutation Prevents Flavonoid Synthesis and Alters Auxin Transport and the Response of Arabidopsis Roots to Gravity and Light[W]. Plant Cell. 2004 May 12;16(5):1191–205.
6. Mo Y, Nagel C, Taylor LP. Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. Proc Natl Acad Sci. 1992 Aug;89(15):7213–7.
7. Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol. 2002 Jun;5(3):218–23.
8. Asen S, Stewart RN, Norris KH. Co-pigmentation of anthocyanins in plant tissues and its effect on color. Phytochemistry. 1972 Mar;11(3):1139–44.
9. Asen S, Stewart RN, Norris KH. Anthocyanin, flavonol copigments, and pH responsible for larkspur flower color. Phytochemistry. 1975 Dec;14(12):2677–82.
10. Emiliani J, Grotewold E, Falcone Ferreyra ML, Casati P. Flavonols protect Arabidopsis plants against UV-B deleterious effects. Mol Plant. 2013 Jul;6(4):1376–9.
11. Jiang N, Doseff AI, Grotewold E. Flavones: From Biosynthesis to Health Benefits. Plants. 2016 Jun;5(2):27.
12. Dixon RA, Sarnala S. Proanthocyanidin Biosynthesis—a Matter of Protection. Plant Physiol. 2020 Oct 1;184(2):579–91.
13. Todd JJ, Vodkin LO. Pigmented Soybean (Glycine max) Seed Coats Accumulate Proanthocyanidins during Development. Plant Physiol. 1993 Jun 1;102(2):663–70.
14. 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 Feb;77(3):367–79.
15. Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK. Natural variation in flavonol and anthocyanin metabolism during cold acclimation in *A rabidopsis thaliana* accessions: Flavonoids and cold acclimation. Plant Cell Environ. 2015 Aug;38(8):1658–72.

16. Becker C, Klaering HP, Kroh LW, Krumbein A. Cool-cultivated red leaf lettuce accumulates cyanidin-3-O-(6″-O-malonyl)-glucoside and caffeoylmalic acid. Food Chem. 2014 Mar;146:404–11.
17. Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK. Flavonoids are determinants of freezing tolerance and cold acclimation in Arabidopsis thaliana. Sci Rep. 2016 Sep 23;6(1):34027.
18. SU W, Zhang G, LI X, GU F, Shi B. Effect of light intensity and light quality on growth and total flavonoid accumulation of Erigeron breviscapus. Chin Tradit Herb Drugs. 1994;
19. Idris A, Linatoc AC, Bakar MFA, Ibrahim ZT, Audu Y. Effect of light quality and quantity on the accumulation of flavonoid in plant species. J Sci Technol. 2018;10(3).
20. Feild TS, Lee DW, Holbrook NM. Why Leaves Turn Red in Autumn. The Role of Anthocyanins in Senescing Leaves of Red-Osier Dogwood. Plant Physiol. 2001 Oct 1;127(2):566–74.
21. falcone ferreyra M, Serra P, Casati P. Recent advances on the roles of flavonoids as plant protective molecules after UV and high light exposure. Physiol Plant. 2021 Aug 1;173.
22. Hanasaki Y, Ogawa S, Fukui S. The correlation between active oxygens scavenging and antioxidative effects of flavonoids. Free Radic Biol Med. 1994 Jun 1;16(6):845–50.
23. Agati G, Azzarello E, Pollastri S, Tattini M. Flavonoids as antioxidants in plants: Location and functional significance. Plant Sci. 2012 Nov 1;196:67–76.
24. Michalak A. Phenolic Compounds and Their Antioxidant Activity in Plants Growing under Heavy Metal Stress. Pol J Environ Stud. 2006 Jan 1;15:523–30.
25. Samanta A, Das G, Das S. Roles of flavonoids in Plants. Int J Pharm Sci Technol. 2011 Jan 1;6:12–35.
26. Subramanian S, Stacey G, Yu O. Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci. 2007 Jul;12(7):282–5.
27. Liu CW, Murray JD. The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. Plants Basel Switz. 2016 Aug 11;5(3):33.
28. Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, et al. Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis. Plant Physiol. 2001 Jun;126(2):524–35.
29. Besseau S, Hoffmann L, Geoffroy P, Catherine L, Pollet B, Legrand M. Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. Plant Cell. 2007 Feb 1;19:148–62.
30. Halbwirth H. The Creation and Physiological Relevance of Divergent Hydroxylation Patterns in the Flavonoid Pathway. Int J Mol Sci. 2010 Feb 4;11(2):595–621.
31. Yoder JI, Belzile F, Tong Y, Goldsbrough A. Visual markers for tomato derived from the anthocyanin biosynthetic pathway. Euphytica. 1994 Jan;79(3):163–7.
32. Shao D, Li Y, Zhu Q, Zhang X, Liu F, Xue F, et al. GhGSTF12, a glutathione S-transferase gene, is essential for anthocyanin accumulation in cotton (Gossypium hirsutum L.). Plant Sci. 2021 Apr 1;305:110827.
33. 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 Oct;6(10):927–38.
34. Zhao ZC, Hu GB, Hu FC, Wang HC, Yang ZY, Lai B. The UDP glucose: flavonoid-3-O-glucosyltransferase (UFGT) gene regulates anthocyanin biosynthesis in litchi (Litchi chinesis Sonn.) during fruit coloration. Mol Biol Rep. 2012 Jun;39(6):6409–15.
35. Francis FJ. Food colorants: anthocyanins. Crit Rev Food Sci Nutr. 1989;28(4):273–314.
36. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. Plant J Cell Mol Biol. 2008 Jun 1;54:733–49.
37. 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.
38. Holton TA, Brugliera F, Tanaka Y. Cloning and expression of flavonol synthase from Petunia hybrida. Plant J. 1993;4(6):1003–10.
39. Davies K, Winefield C, Lewis D, Nielsen K, Bradley M, Schwinn K, et al. Research into control of flower colour and flowering time in Eustoma grandiflorum (Lisanthus). Flower Newsl. 1997 Jan 1;23:24–32.
40. 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 [Internet]. 2016 [cited 2023 Jul 7];6. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2015.01257
41. Nielsen K, Deroles SC, Markham KR, Bradley MJ, Podivinsky E, Manson D. Antisense flavonol synthase alters copigmentation and flower color in lisianthus. Mol Breed. 2002 Dec 1;9(4):217–29.
42. Tanaka Y, Brugliera F. Flower colour and cytochromes P450. Philos Trans R Soc B Biol Sci. 2013 Feb 19;368(1612):20120432.
43. Halbwirth H, Miosic S, Milosevic M, Nitarska D, Thill J, Stich K, et al. Re-investigating substrate specificity of dihydroflavonol 4-reductase with respect to the B-ring hydroxylation pattern of substrates. Acta Hortic. 2019 Jul;(1242):889–98.
44. 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: TFs of A. thaliana R2R3-MYB subgroup 7. Plant J. 2007 Apr 25;50(4):660–77.
45. Schilbert HM, Glover BJ. Analysis of flavonol regulator evolution in the Brassicaceae reveals MYB12, MYB111 and MYB21 duplications and MYB11 and MYB24 gene loss. BMC Genomics. 2022 Aug 19;23(1):604.
46. Battat M, Eitan A, Rogachev I, Hanhineva K, Fernie A, Tohge T, et al. A MYB Triad Controls Primary and Phenylpropanoid Metabolites for Pollen Coat Patterning. Plant Physiol. 2019 May;180(1):87–108.
47. Zhang X, He Y, Li L, Liu H, Hong G. Involvement of the R2R3-MYB transcription factor MYB21 and its homologs in regulating flavonol accumulation in Arabidopsis stamen. Hancock R, editor. J Exp Bot. 2021 May 28;72(12):4319–32.
48. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci. 2015 Mar;20(3):176–85.
49. LaFountain AM, Yuan Y. Repressors of anthocyanin biosynthesis. New Phytol. 2021 Aug;231(3):933–49.
50. Johnson ET, Ryu S, Yi H, Shin B, Cheong H, Choi G. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase: Substrate specificity of DFR. Plant J. 2001 Dec 23;25(3):325–33.
51. Forkmann G, Ruhnau B. Distinct Substrate Specificity of Dihydroflavonol 4-Reductase from Flowers of Petunia hybrida. Z Für Naturforschung C. 1987 Oct 1;42(9–10):1146–8.
52. 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 [Internet]. 2017 Mar 28 [cited 2023 Jul 7];8. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2017.00428/full
53. Gerats AGM, De Vlaming P, Doodeman M, Al B, Schram AW. Genetic control of the conversion of dihydroflavonols into flavonols and anthocyanins in flowers of Petunia hybrida. Planta. 1982 Aug;155(4):364–8.
54. Beld M, Martin C, Huits H, Stuitje AR, Gerats AGM. Flavonoid synthesis in Petunia hybrida: partial characterization of dihydroflavonol-4-reductase genes. Plant Mol Biol. 1989 Nov;13(5):491–502.
55. Helariutta Y, Elomaa P, Kotilainen M, Griesbach RJ, Schr❖der J, Teeri TH. Chalcone

synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in Gerbera hybrida (Asteraceae). Plant Mol Biol. 1995 Apr;28(1):47–60.
56. Xie DY, Jackson LA, Cooper JD, Ferreira D, Paiva NL. Molecular and Biochemical Analysis of Two cDNA Clones Encoding Dihydroflavonol-4-Reductase from *Medicago truncatula*. Plant Physiol. 2004 Mar 1;134(3):979–94.
57. Shimada N, Sasaki R, Sato S, Kaneko T, Tabata S, Aoki T, et al. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the Lotus japonicus genome. J Exp Bot. 2005 Sep 1;56(419):2573–85.
58. Petit P, Granier T, d'Estaintot BL, Manigand C, Bathany K, Schmitter JM, et al. Crystal Structure of Grape Dihydroflavonol 4-Reductase, a Key Enzyme in Flavonoid Biosynthesis. J Mol Biol. 2007 May 18;368(5):1345–57.
59. Park S, Kim DH, Park BR, Lee JY, Lim SH. Molecular and Functional Characterization of *Oryza sativa* Flavonol Synthase (OsFLS), a Bifunctional Dioxygenase. J Agric Food Chem. 2019 Jul 3;67(26):7399–409.
60. Falcone Ferreyra ML, Rius S, Emiliani J, Pourcel L, Feller A, Morohashi K, et al. Cloning and characterization of a UV-B-inducible maize flavonol synthase. Plant J. 2010;62(1):77–91.
61. Owens DK, Alerding AB, Crosby KC, Bandara AB, Westwood JH, Winkel BSJ. Functional Analysis of a Predicted Flavonol Synthase Gene Family in Arabidopsis. Plant Physiol. 2008 Jul 8;147(3):1046–61.
62. Wellmann F, Lukačin R, Moriguchi T, Britsch L, Schiltz E, Matern U. Functional expression and mutational analysis of flavonol synthase from *Citrus unshiu*: Flavonol synthase. Eur J Biochem. 2002 Aug;269(16):4134–42.
63. Jiang X, Shi Y, Fu Z, Li WW, Lai S, Wu Y, et al. Functional characterization of three flavonol synthase genes from Camellia sinensis: Roles in flavonol accumulation. Plant Sci. 2020 Nov;300:110632.
64. Park S, Kim DH, Lee JY, Ha SH, Lim SH. Comparative Analysis of Two Flavonol Synthases from Different-Colored Onions Provides Insight into Flavonoid Biosynthesis. J Agric Food Chem. 2017 Jul 5;65(26):5287–98.
65. Li C, Bai Y, Li S, Chen H, Han X, Zhao H, et al. Cloning, Characterization, and Activity Analysis of a Flavonol Synthase Gene *FtFLS1* and Its Association with Flavonoid Content in Tartary Buckwheat. J Agric Food Chem. 2012 May 23;60(20):5161–8.
66. Kou M, Li C, Song W, Shen Y, Tang W, Zhang Y, et al. Identification and functional characterization of a flavonol synthase gene from sweet potato [Ipomoea batatas (L.) Lam.]. Front Plant Sci. 2023 May 10;14:1181173.
67. Shi J, Li W, Gao Y, Wang B, Li Y, Song Z. Enhanced rutin accumulation in tobacco leaves by overexpressing the *NtFLS2* gene. Biosci Biotechnol Biochem. 2017 Sep 2;81(9):1721–5.
68. Chua CS, Biermann D, Goo KS, Sim TS. Elucidation of active site residues of Arabidopsis thaliana flavonol synthase provides a molecular platform for engineering flavonols. Phytochemistry. 2008 Jan;69(1):66–75.
69. Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield CJ. Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis thaliana. Structure. 2002 Jan;10(1):93–103.
70. Østergaard L, Lauvergeat V, Næsted H, Mattsson O, Mundy J. Two differentially regulated Arabidopsis genes define a new branch of the DFR superfamily. Plant Sci. 2001 Feb;160(3):463–72.
71. Katsu K, Suzuki R, Tsuchiya W, Inagaki N, Yamazaki T, Hisano T, et al. A new buckwheat dihydroflavonol 4-reductase (DFR), with a unique substrate binding structure, has altered substrate specificity. BMC Plant Biol. 2017 Dec;17(1):239.
72. Miosic S, Thill J, Milosevic M, Gosch C, Pober S, Molitor C, et al. Dihydroflavonol

4-Reductase Genes Encode Enzymes with Contrasting Substrate Specificity and Show Divergent Gene Expression Profiles in Fragaria Species. PLOS ONE. 2014 Nov 13;9(11):e112707.
73. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012 Jan;40(D1):D1178–86.
74. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2016 Jan 4;44(Database issue):D67–72.
75. Pucker B. script_collection [Internet]. 2023 [cited 2023 Jul 21]. Available from: https://github.com/bpucker/script_collection
76. 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.
77. Rempel A, Choudhary N, Pucker B. KIPEs3: Automatic annotation of biosynthesis pathways [Internet]. Bioinformatics; 2022 Jul [cited 2023 Jul 21]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.06.30.498365
78. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr;30(4):772–80.
79. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992 Nov 15;89(22):10915–9.
80. Brown JW, Walker JF, Smith SA. Phyx: phylogenetic tools for unix. Kelso J, editor. Bioinformatics. 2017 Jun 15;33(12):1886–8.
81. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587–9.
82. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Teeling E, editor. Mol Biol Evol. 2020 May 1;37(5):1530–4.
83. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. Poon AFY, editor. PLoS ONE. 2010 Mar 10;5(3):e9490.
84. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Battistuzzi FU, editor. Mol Biol Evol. 2021 Jun 25;38(7):3022–7.
85. 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.
86. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. Nucleic Acids Res. 2011 Jan;39(Database issue):D19–21.
87. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 May;34(5):525–7.
88. Pucker B, Iorizzo M. Apiaceae FNS I originated from F3H through tandem gene duplication. Remington DL, editor. PLOS ONE. 2023 Jan 19;18(1):e0280155.
89. PBBtools/collection/generate_subset.py at main · bpucker/PBBtools · GitHub [Internet]. [cited 2023 Oct 6]. Available from: https://github.com/bpucker/PBBtools/blob/main/collection/generate_subset.py
90. 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 [Internet]. Evolutionary Biology; 2022 Oct [cited 2023 Jul 21]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.10.19.512958
91. Waskom M. seaborn: statistical data visualization. J Open Source Softw. 2021 Apr 6;6(60):3021.
92. Stracke R, De Vos RCH, Bartelniewoehner L, Ishihara H, Sagasser M, Martens S, et al. Metabolomic and genetic analyses of flavonol synthesis in Arabidopsis thaliana support the in

vivo involvement of leucoanthocyanidin dioxygenase. Planta. 2009 Jan 1;229(2):427–45.
93. Schilbert HM, Schöne M, Baier T, Busche M, Viehöver P, Weisshaar B, et al. Characterization of the Brassica napus Flavonol Synthase Gene Family Reveals Bifunctional Flavonol Synthases. Front Plant Sci [Internet]. 2021 [cited 2023 Oct 20];12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.733762
94. The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016 May 1;181(1):1–20.
95. Gebhardt Y, Witte S, Forkmann G, Lukačin R, Matern U, Martens S. Molecular evolution of flavonoid dioxygenases in the family Apiaceae. Phytochemistry. 2005 Jun 1;66(11):1273–84.
96. Johnson ET, Yi H, Shin B, Oh BJ, Cheong H, Choi G. Cymbidium hybrida dihydroflavonol 4-reductase does not efficiently reduce dihydrokaempferol to produce orange pelargonidin-type anthocyanins. Plant J. 1999;19(1):81–5.
97. Žilić S, Serpen A, Akıllıoğlu G, Gökmen V, Vančetović J. Phenolic Compounds, Carotenoids, Anthocyanins, and Antioxidant Capacity of Colored Maize (Zea mays L.) Kernels. J Agric Food Chem. 2012 Feb 8;60(5):1224–31.
98. Aoki H, Kuze N, Kato Y, Ei S, Inc F. Anthocyanin isolated from purple corn (Zea mays L). Foods Food Ingred J Jpn. 2001 Jan 1;199.
99. Jende-Strid B. Gene-enzyme relations in the pathway of flavonoid biosynthesis in barley. Theor Appl Genet. 1991 May 1;81(5):668–74.
100. Nyegaard Kristiansen K, Rohde W. Structure of the Hordeum vulgare gene encoding dihydroflavonol-4-reductase and molecular analysis of ant18 mutants blocked in flavonoid synthesis. Mol Gen Genet MGG. 1991 Nov 1;230(1):49–59.
101. Liu H, Lou Q, Ma J, Su B, Gao Z, Liu Y. Cloning and Functional Characterization of Dihydroflavonol 4-Reductase Gene Involved in Anthocyanidin Biosynthesis of Grape Hyacinth. Int J Mol Sci. 2019 Jan;20(19):4743.
102. Leonard E, Yan Y, Chemler J, Matern U, Martens S, Koffas MAG. Characterization of dihydroflavonol 4-reductases for recombinant plant pigment biosynthesis applications. Biocatal Biotransformation. 2008 Jan 1;26(3):243–51.
103. Almeida JRM, D'Amico E, Preuss A, Carbone F, de Vos CHR, Deiml B, et al. Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (Fragaria ×ananassa). Arch Biochem Biophys. 2007 Sep 1;465(1):61–71.
104. Li P, Lei K, Liu L, Zhang G, Ge H, Zheng C, et al. Identification and functional characterization of a new flavonoid synthase gene MdFLS1 from apple. Planta. 2021 Apr 15;253(5):105.
105. Suzuki K ichi, Tsuda Sh, Fukui Y, Fukuchi-Mizutani M, Yonekura-Sakakibara K, Tanaka Y, et al. Molecular Characterization of Rose Flavonoid Biosynthesis Genes and Their Application in Petunia. Biotechnol Biotechnol Equip. 2000 Jan 1;14(2):56–62.
106. Fischer TC, Halbwirth H, Meisel B, Stich K, Forkmann G. Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from Malus domestica and Pyrus communis cultivars and the consequences for flavonoid metabolism. Arch Biochem Biophys. 2003 Apr 15;412(2):223–30.
107. Tanaka Y, Fukui Y, Fukuchi-Mizutani M, Holton TA, Higgins E, Kusumi T. Molecular Cloning and Characterization of Rosa hybrida Dihydroflavonol 4-reductase Gene. Plant Cell Physiol. 1995 Sep 1;36(6):1023–31.
108. Spribille R, Forkmann G. Conversion of Dihydroflavonols to Flavonols with Enzyme Extracts from Flower Buds of Matthiola incana R. Br. Z Für Naturforschung C. 1984 Aug 1;39(7–8):714–9.
109. Turnbull JJ, Nakajima J ichiro, Welford RWD, Yamazaki M, Saito K, Schofield CJ.

Mechanistic Studies on Three 2-Oxoglutarate-dependent Oxygenases of Flavonoid Biosynthesis: ANTHOCYANIDIN SYNTHASE, FLAVONOL SYNTHASE, AND FLAVANONE 3β-HYDROXYLASE*. J Biol Chem. 2004 Jan 9;279(2):1206–16.
110. Bloor SJ, Abrahams S. The structure of the major anthocyanin in Arabidopsis thaliana. Phytochemistry. 2002 Feb 1;59(3):343–6.
111. Martens S, Teeri T, Forkmann G. Heterologous expression of dihydroflavonol 4-reductases from various plants. FEBS Lett. 2002 Nov 20;531(3):453–8.
112. Saito N, Tatsuzawa F, Nishiyama A, Yokoi M, Shigihara A, Honda T. Acylated cyanidin 3-sambubioside-5-glucosides in Matthiola incana. Phytochemistry. 1995 Mar;38(4):1027–32.
113. Aiguo Z, Ruiwen D, Cheng W, Cheng C, Dongmei W. Insights into the catalytic and regulatory mechanisms of dihydroflavonol 4-reductase, a key enzyme of anthocyanin synthesis in Zanthoxylum bungeanum. Tree Physiol. 2023 Jan 1;43(1):169–84.
114. Stich K, Eidenberger T, Wurst F, Forkmann G. Flavonol Synthase Activity and the Regulation of Flavonol and Anthocyanin Biosynthesis during Flower Development in Dianthus caryophyllus L. (Carnation). Z Für Naturforschung C. 1992 Aug 1;47(7–8):553–60.
115. Jiang J. Cloning of flavonol synthase gene from Fagopyum dibotrys and its expression in Escherichia coli. Chin Tradit Herb Drugs. 2013;1974–8.
116. McCarthy EW, Landis JB, McCoy AG, Lawhorn AJ, Kurti A, Xu Y, et al. Homeologue differential expression in the flavonoid biosynthetic pathway underlies flower colour variation in natural and synthetic polyploids of Nicotiana tabacum (Solanaceae). Bot J Linn Soc. 2023 Oct 24;boad052.
117. Tanaka Y, Yonekura K, Fukuchi-Mizutani M, Fukui Y, Fujiwara H, Ashikari T, et al. Molecular and Biochemical Characterization of Three Anthocyanin Synthetic Enzymes from Gentiana triflora. Plant Cell Physiol. 1996 Jul 1;37(5):711–6.
118. Gottlieb OR. Micromolecular Evolution, Systematics and Ecology: An Essay into a Novel Botanical Discipline. Springer Science & Business Media; 2012. 183 p.
119. Harborne JB, Williams CA. Advances in flavonoid research since 1992. Phytochemistry. 2000 Nov 1;55(6):481–504.
120. Xu F, Cheng H, Cai R, Li LL, Chang J, Zhu J, et al. Molecular Cloning and Function Analysis of an Anthocyanidin Synthase Gene from Ginkgo biloba, and Its Expression in Abiotic Stress Responses. Mol Cells. 2008 Dec 31;26(6):536–47.
121. Reddy AM, Reddy VS, Scheffler BE, Wienand U, Reddy AR. Novel transgenic rice overexpressing anthocyanidin synthase accumulates a mixture of flavonoids leading to an increased antioxidant potential. Metab Eng. 2007 Jan 1;9(1):95–111.
122. Pang Y, Peel GJ, Wright E, Wang Z, Dixon RA. Early Steps in Proanthocyanidin Biosynthesis in the Model Legume Medicago truncatula. Plant Physiol. 2007 Nov;145(3):601–15.
123. Martens S, Preuß A, Matern U. Multifunctional flavonoid dioxygenases: Flavonol and anthocyanin biosynthesis in Arabidopsis thaliana L. Phytochemistry. 2010 Jul 1;71(10):1040–9.
124. Kawai Y, Ono E, Mizutani M. Evolution and diversity of the 2–oxoglutarate-dependent dioxygenase superfamily in plants. Plant J. 2014;78(2):328–43.
125. Wang Z, Wang S, Wu M, Li Z, Liu P, Li F, et al. Evolutionary and functional analyses of the 2-oxoglutarate-dependent dioxygenase genes involved in the flavonoid biosynthesis pathway in tobacco. Planta. 2019 Feb 1;249(2):543–61.
126. Preuss A, Stracke R, Weisshaar B, Hillebrecht A, Matern U, Martens S. Arabidopsis thaliana expresses a second functional flavonol synthase. FEBS Lett. 2009 Jun 18;583(12):1981–6.
127. Moriguchi T, Kita M, Ogawa K, Tomono Y, Endo T, Omura M. Flavonol synthase gene expression during citrus fruit development. Physiol Plant. 2002;114(2):251–8.
128. FUJITA A, GOTO-YAMAMOTO N, ARAMAKI I, HASHIZUME K. Organ-Specific

Transcription of Putative Flavonol Synthase Genes of Grapevine and Effects of Plant Hormones and Shading on Flavonol Biosynthesis in Grape Berry Skins. Biosci Biotechnol Biochem. 2006 Jan 1;70(3):632–8.
129. Pucker B, Choudhary N. Gene expression data sets of selected plant species. [cited 2023 Jul 21]; Available from: https://leopard.tu-braunschweig.dereceive/dbbs_mods_00073263

Additional Files

Additional File 1: References of the coding sequence data sets that were used in this study.
Additional File 2: Collection of ANR, CCR, DFR and DFR-like sequences that were used for the phylogenetic analysis of DFR.
Additional File 3: Collection of F3H, FNS1, ANS, FLS and FLS-like sequences that were used for the phylogenetic analyses of FLS.
Additional File 4: DFR phylogenetic trees. (1) DFR tree showing outgroups, (2a-f) DFR-specific trees when the outgroups are collapsed. (2a) Constructed by IQ-TREE based on a MAFFT alignment, (2b) constructed by IQ-TREE based on a Muscle5 alignment, (2c) constructed by FastTree based on a MAFFT alignment, (2d) constructed by FastTree based on a Muscle5 alignment, (2e) constructed by MEGA based on a MAFFT alignment, and (2f) constructed by MEGA based on a Muscle5 alignment.
Additional File 5: References of previously characterised FLS and DFR included in phylogenetic tree analyses.
Additional File 6: FLS phylogenetic trees. (1) FLS tree showing outgroups, (2a-f) FLS-specific trees when the outgroups are collapsed. (2a) Constructed by IQ-TREE based on a MAFFT alignment, (2b) constructed by IQ-TREE based on a Muscle5 alignment, (2c) constructed by fasttree based on a MAFFT alignment, (2d) constructed by FastTree based on a Muscle5 alignment, (2e) constructed by MEGA based on a MAFFT alignment, and (2f) constructed by MEGA based on a Muscle5 alignment.
Additional File 7: Extended Heatmap of 43 species depicting the transcript abundance of analysed genes in different tissues.