

7-25-2019

# Assessing Anthocyanin Biosynthesis in Solanaceae as a Model Pathway for Secondary Metabolism

Zou Li

*University of Nebraska - Lincoln & Guangdong Academy of Agricultural Sciences*

Trisha Vickrey

*University of Nebraska-Lincoln, tvickrey2@unl.edu*

Moira G. McNally

*University of Nebraska - Lincoln*

Shirley J. Sato


*University of Nebraska - Lincoln*

Thomas E. Clemente

*University of Nebraska - Lincoln, tclemente1@unl.edu*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/plantscifacpub>

 Part of the [Plant Biology Commons](#), [Plant Breeding and Genetics Commons](#), and the [Plant Pathology Commons](#)

Li, Zou; Vickrey, Trisha; McNally, Moira G.; Sato, Shirley J.; Clemente, Thomas E.; and Mower, Jeffrey P., "Assessing Anthocyanin Biosynthesis in Solanaceae as a Model Pathway for Secondary Metabolism" (2019). *Faculty Publications from the Center for Plant Science Innovation*. 210.

<https://digitalcommons.unl.edu/plantscifacpub/210>

This Article is brought to you for free and open access by the Plant Science Innovation, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Center for Plant Science Innovation by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Zou Li, Trisha Vickrey, Moira G. McNally, Shirley J. Sato, Thomas E. Clemente, and Jeffrey P. Mower

Article

# Assessing Anthocyanin Biosynthesis in Solanaceae as a Model Pathway for Secondary Metabolism

Zuo Li <sup>1,2,†</sup>, Trisha L. Vickrey <sup>1,3,†</sup>, Moira G. McNally <sup>1,4</sup>, Shirley J. Sato <sup>1,5</sup>,  
Tom Elmo Clemente <sup>1,6</sup> and Jeffrey P. Mower <sup>1,6,\*</sup>

<sup>1</sup> Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA

<sup>2</sup> Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

<sup>3</sup> Department of Chemistry, University of Nebraska, Lincoln, NE 68588, USA

<sup>4</sup> Biology Department, University of Jamestown, Jamestown, ND 58405, USA

<sup>5</sup> Center for Biotechnology, University of Nebraska, Lincoln, NE 68588, USA

<sup>6</sup> Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

\* Correspondence: [jpmower@unl.edu](mailto:jpmower@unl.edu)

† These authors contributed equally to this work.

Received: 19 June 2019; Accepted: 23 July 2019; Published: 25 July 2019



**Abstract:** Solanaceae have played an important role in elucidating how flower color is specified by the flavonoid biosynthesis pathway (FBP), which produces anthocyanins and other secondary metabolites. With well-established reverse genetics tools and rich genomic resources, Solanaceae provide a robust framework to examine the diversification of this well-studied pathway over short evolutionary timescales and to evaluate the predictability of genetic perturbation on pathway flux. Genomes of eight Solanaceae species, nine related asterids, and four rosids were mined to evaluate variation in copy number of the suite of FBP enzymes involved in anthocyanin biosynthesis. Comparison of annotation sources indicated that the NCBI annotation pipeline generated more and longer FBP annotations on average than genome-specific annotation pipelines. The pattern of diversification of each enzyme among asterids was assessed by phylogenetic analysis, showing that the CHS superfamily encompasses a large paralogous family of ancient and recent duplicates, whereas other FBP enzymes have diversified via recent duplications in particular lineages. Heterologous expression of a pansy F3'5'H gene in tobacco changed flower color from pink to dark purple, demonstrating that anthocyanin production can be predictably modified using reverse genetics. These results suggest that the Solanaceae FBP could be an ideal system to model genotype-to-phenotype interactions for secondary metabolism.

**Keywords:** anthocyanin biosynthesis; comparative genomics; reverse genetics; Solanaceae

## 1. Introduction

In many plants, anthocyanins are the major secondary metabolites that determine fruit and flower color, which is important for plant reproduction by attracting agents for pollination and seed dispersal [1]. The production of anthocyanidins, the immediate precursors of anthocyanins, involves a series of enzymatic reactions in the flavonoid biosynthesis pathway (FBP). Chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone 3-hydroxylase (F3H) are early pathway enzymes that commit to the production of flavonoids. Flavonoid 3'-monooxygenase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) direct flux to a particular branch of the pathway to produce different anthocyanidins with distinct color profiles, and dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS) make additional modifications to generate three major anthocyanidins

(cyanidin, delphinidin, pelargonidin), from which three additional anthocyanidins (malvidin, peonidin, petunidin) can be derived by methylation. Expression of FBP enzymes is regulated by a suite of transcription factors in the R2R3-MYB, bHLH, and WDR multigene families [2,3]. In addition to anthocyanins, the FBP produces an array of other plant secondary compounds (e.g., flavones, flavonols, isoflavonoids, lignins, and tannins) that participate in plant ecological and physiological functions, such as stress protection, auxin transport and nodulation signaling [1,4–6].

Solanaceae (the nightshade family) are an emerging powerhouse for plant evolutionary and comparative genomics. This family comprises 90–100 genera and 2500–3000 species [7,8], including many economically important plants such as tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), tomato (*S. lycopersicum*), petunia (*Petunia hybrida*), sweet and spicy peppers (*Capsicum* spp.), and eggplant (*S. melongena*). To date, over a dozen draft genomes are available from these and other major solanaceous crops as well as many of their progenitors and wild relatives [9–21]. In addition, genomes from successively more distantly related species of asterids are available, providing broader evolutionary context for comparative analyses. The morning glory (*Ipomoea nil* [22]) genome represents Convolvulaceae, the sister family to Solanaceae in Solanales. Lamiales and Gentianales, the two orders closest to Solanales, are represented by genomes from monkeyflower (*Mimulus guttatus* [23]), sesame (*Sesamum indicum* [24]), robusta coffee (*Coffea canephora* [25]), and olive (*Olea europaea* [26,27]). Sister to the lamiids, to which all aforementioned plants belong, are the campanulids, with genomes available for carrot (*Daucus carota* [28]), sunflower (*Helianthus annuus* [29]), lettuce (*Lactuca sativa* [30]), and wild artichoke (*Cynara cardunculus* [31]).

With such diverse availability of genomes from a single family, Solanaceae provide an ideal platform to perform comprehensive studies on the evolutionary diversification of pathways such as the FBP. Indeed, many Solanaceae species exhibit diversity in the types and relative levels of different anthocyanins in their flowers and fruits: pink tobacco flowers primarily contain cyanidin [32], the dark purple color of eggplant skins and *Ipomoea* flowers is produced by delphinidin [33,34], many red-fleshed potato cultivars derive their coloration from pelargonidin [35,36], and all six of the major anthocyanidins can be variously found in different potato cultivars [37]. Moreover, petunia and tobacco have played key roles in the elucidation of FBP enzyme function due to the early availability of efficient plant transformation systems for these species. Transgenic tobacco and petunia lines have been used to demonstrate increased anthocyanin biosynthesis by introduction or overexpression of *DFR*, *F3'H*, or *F3'5'H* [38–41], reduced anthocyanin biosynthesis by antisense RNA or RNAi suppression of *CHS* or *CHI* [42–44], and altered FBP flux after introduction or suppression of FBP enzymes that catalyze the production of non-anthocyanin products such as flavonols, flavones, isoflavones, and stilbenoids [41,45–48].

Here, we used the extensive genomic and reverse genetic resources for Solanaceae and related species to advance anthocyanin biosynthesis in Solanaceae as a model pathway for secondary metabolism. We conducted a comprehensive survey on the diversity of FBP enzymes involved in anthocyanin biosynthesis to establish the variation in presence and copy-number of anthocyanin biosynthesis genes among Solanaceae and other eudicots. Because gene copy number is dependent on the quality of gene annotation, we also assessed the consistency of FBP gene annotation between different annotation sources for most species. Phylogenetic analysis was used to identify the pattern of FBP enzyme diversification via gene duplication. Finally, a useful system for pathway modeling requires that simple pathway perturbations should result in predictable changes in phenotype. Thus, we transgenically introduced pansy *F3'5'H* into tobacco to demonstrate the reliability with which flower color can be predictably altered via genetic modification. These results contribute to the extensive body of literature on the evolution and genetic manipulation of the FBP pathway in Solanaceae, and sets up the Solanaceae FBP as a potentially powerful system to model genotype-to-phenotype interactions in secondary metabolism.

## 2. Materials and Methods

### 2.1. Identification of Flavonoid Biosynthesis Pathway Homologs

Gene sequences from six representative eudicots (*Arabidopsis thaliana*, *D. carota*, *M. guttatus*, *S. lycopersicum*, *S. tuberosum*, and *Vitis vinifera*) were downloaded from Phytozome using KEGG codes for CHS (KEGG code K00660), CHI (K01859), F3H (K00475), F3'H (K05280), F3'5'H (K13083), DFR (K13082), and ANS (K05277). These sequences were used as queries in relaxed blastn searches (-task blastn -penalty -1 -reward 1 -evaluate  $1e^{-50}$ ) to find coding sequence homologs in 39 independent genome annotations from 21 species of eudicots, with a focus on solanaceous crops. Annotated coding sequences created by the NCBI Eukaryotic Genome Annotation Pipeline were obtained for 18 species, while annotated coding sequences created by genome-specific pipelines were obtained from Phytozome version 12.1 for 10 species, from the Sol Genomics Network for 8 species, and from genome-specific websites for 3 species (Table S1). In total, two independent sets of annotated coding sequences were obtained for 18 species, while a single annotation was obtained for 3 species.

When homologous sequences included 5' or 3' UTRs, the program TransDecoder v5.3 [49] was used to infer the coding sequence, requiring a minimum coding sequence length of 300 bp. For each species, transcript variants arising from a single locus were eliminated by several approaches, depending on source of the annotated gene sequences. For annotations in Phytozome, the 'primary transcripts only' data sets were used, which are prefiltered to contain a single transcript per locus. For NCBI pipeline annotations, transcript variants were excluded by keeping the longest transcript sequence for every defined locus number. For data downloaded from the Sol Genomics Network and species-specific websites, generic feature format annotation files were consulted to keep only the single longest annotated transcript for each locus.

### 2.2. Comparison of Annotations from Different Sources

To assess the level of agreement between annotations from different sources, annotations from the NCBI Eukaryotic Genome Annotation Pipeline were compared with annotations from genome-specific pipelines (available from Phytozome, Sol Genomics Network, and genomic-specific websites) for 18 species (Table S1). Using blastn with default values, the annotated genes from each source were classified as perfect pairs, imperfect pairs, or unpaired. 'Perfect pairs' were reciprocal-best hits (RBHs) with 100% identity over 100% of the length of the query sequence in a default blastn search. 'Imperfect pairs' were RBHs with <100% identity and/or different length relative to the query sequence. Annotations that did not have a perfect or imperfect pair between sources were initially classified as 'unpaired'. Unpaired annotations were excluded if their coding sequences were identical to another annotated gene from the same annotation source.

### 2.3. Phylogenetic Analysis of Flavonoid Biosynthesis Enzymes

Gene annotations from both annotation sources were examined phylogenetically to determine whether any unpaired genes clustered in a phylogenetic tree. DNA sequences for each flavonoid biosynthesis enzyme were aligned with Muscle v3.8.31 [50] using default parameters. Aligned enzyme sequences were trimmed using Gblocks v0.91b [51] in codon mode, with the minimum block length parameter (b4) set to 5 and the parameters for minimum flanking position sequence (b2) and gaps allowed (b5) set to half the total number of aligned sequences. Phylogenetic trees were constructed from each of the trimmed sequence alignments using the maximum likelihood implementation in RAxML v8.2.4 [52]. For each enzyme data set, the general time reversible model with a gamma distribution of rate variation was used. During the analysis, the shape of the gamma distribution and substitution rate parameters were estimated, and a fast bootstrap analysis with 1000 replicates was performed to assess support for branches. Unpaired genes from different annotation sources that formed a unique cluster in the phylogenetic results were reclassified as 'imperfect pairs'.

Final data sets were created for each FBP gene by keeping only one annotation of the perfect and imperfect pairs, and in these cases, the NCBI pipeline annotations were kept over the genome-specific pipeline annotations for consistency. These final data sets for each FBP gene were aligned using Muscle, trimmed using Gblocks, and phylogenetically analyzed using RAxML as described above. Because the CHS data set comprises a family of paralogous genes, the resulting tree was rooted using non-asterid species in order to evaluate the relationship among asterid CHS paralogs. All other genes were unrooted analyses of asterid sequences.

#### 2.4. Evaluation of Flower Color Modification Using Transgenic *F3'5'H*

The *F3'5'H* gene from *Viola × wittrockiana*, which was previously shown to modify flower color from pink/red hues to blue/purple hues in several plant species [53,54], was chosen for transgene analysis in this study. An expression cassette was synthesized (GenScript Corp., Piscataway, NJ, USA) containing the complete *V. wittrockiana F3'5'H* coding sequence (GenBank accession number AB332097) coupled with a translational enhancer from tobacco etch virus, which were situated between an enhanced 35S promoter from cauliflower mosaic virus and a polyadenylation signal. This expression cassette was subcloned into the binary vector pPZP212, which contains the selectable marker *NptII* [55], and introduced into *Agrobacterium tumefaciens* C58C1/pMP90 by triparental mating [56]. Transformation was carried out using tobacco (*N. tabacum* cv. Xanthi) leaf discs according to established protocols [57]. Regenerated plants ( $T_0$  generation) and seeds collected from self-pollinated  $T_0$  plants ( $T_1$  generation) or wild type Xanthi plants (WT) were grown to maturity in a glasshouse, and flower color was assessed by eye.

Expression of *NptII* was monitored in selected  $T_0$  and  $T_1$  plants by an NPTII ELISA kit (Agdia Inc., Elkhart, IN, USA) according to the supplied protocol. The presence and expression of the *F3'5'H* transgene was evaluated in selected  $T_0$  and  $T_1$  plants through polymerase chain reaction (PCR) and reverse transcription PCR (RT-PCR), respectively, using the native *RPL23A* gene as a standard control. To assess the presence of the transgene, genomic DNA was extracted from a mature leaf according to a CTAB procedure [58], and PCR was performed with 20 ng of DNA using gene-specific primers for *F3'5'H* (CCGAGTCTAACGAGTTCAAAG; CCCATTTGTATTGTTCGCATTC) and *RPL23A* (GCACCTGGAAGGAACAAA; ACGTCCAAAGCATCATAGTC) according to a previously described PCR protocol [59]. Cellular RNA was extracted from a mature flower and 2  $\mu$ g was converted to cDNA according to established procedures [60]. Expression of *F3'5'H* and *RPL23A* was assayed by RT-PCR using a semi-quantitative approach in which reactions were amplified for 20, 24, 28, or 32 cycles of denaturation (95 °C for 30 s), annealing (55 °C for 30 s), and elongation (72 °C for 90 s).

To compare anthocyanin content of  $T_1$  flowers to WT, the corolla lobe from individual flowers were collected upon full flower opening and then stored at  $-80$  °C. Anthocyanins were eluted overnight in 1 mL of acidified methanol with 1% HCl. Absorbance was measured at 530 nm using a GENESYS 150 UV-Vis spectrophotometer for 3–4 flowers from each of four  $T_1$  transgenic lines and WT. Absorbance values from transgenic lines were compared to WT plants using an unpaired *t*-test. For thin-layer chromatography analysis, silica gel plates (10  $\times$  20 cm, Merck) were used as a stationary phase and n-butanol:acetic acid:water (4:1:5, v/v/v, upper phase) was used as a mobile phase. Approximately 2  $\mu$ L of sample was spotted onto the plate by adding a small amount at a time using a 10  $\mu$ L-pipette. Samples were spaced 1 cm apart. Plates were developed in a tank with the solvent mixture. After developing, the plates were dried and pigments visualized directly.

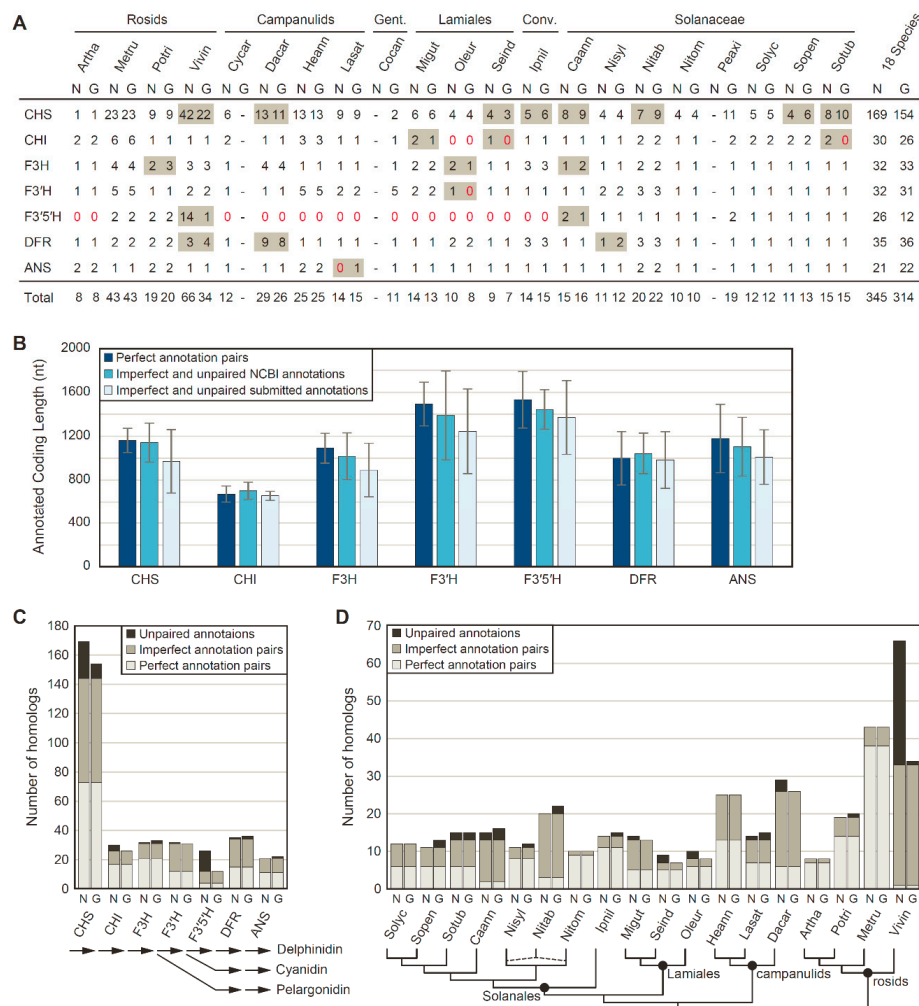
### 3. Results

#### 3.1. The NCBI Pipeline Annotates More and Longer Anthocyanin Biosynthesis Enzymes

Putative homologs to seven FBP enzymes (CHS, CHI, F3H, F3'H, F3'5'H, DFR, and ANS) were identified in each of 39 genome annotations from 21 different eudicots by blast homology searches, followed by filtration to remove short coding sequences (<300 bp) and transcript variants (Figure 1A).



To assess annotation consistency between NCBI pipeline annotations and genome-specific pipeline annotations stored in various data repositories, blast searches and phylogenetic clustering were used to categorize each FBP homolog as a perfect pair, an imperfect pair, or an unpaired sequence for the 18 species with both annotation types. For these 18 species, 345 FBP homologs were identified in the NCBI-based pipeline annotations, whereas 314 FBP homologs were identified in the genome-specific pipeline annotations. Of these, 299 homologs were shared between both annotation sets, with identical coding sequence annotations for about half of the shared homologs (153 perfect pairwise matches; 146 imperfect pairwise matches). However, 46 NCBI pipeline annotations and 15 genome-specific pipeline annotations did not have a pairwise match in the other annotation set. In terms of length, the imperfect pairs and unpaired annotations were longer for all enzymes on average from the NCBI pipeline annotation compared with genome-specific pipeline annotations (Figure 1B).



**Figure 1.** Homology detection of anthocyanin biosynthesis genes between NCBI (N) and genome-specific (G) pipeline annotations. (A) Anthocyanin biosynthesis gene counts. Discrepant values between annotation sources are shaded in gray. Zero values are denoted with red text. The sum totals at right include only the 18 species with NCBI and genome-specific pipeline annotations. Conv., Convolvulaceae; Gent., Gentianales. (B) Mean  $\pm$  SD for coding sequence lengths in nucleotides (nt) of anthocyanin biosynthesis genes. (C) Number of homologs identified for each anthocyanin biosynthesis gene. A simplified anthocyanin biosynthesis pathway is shown at the bottom. (D) Number of homologs identified for each species. Phylogenetic relationships among sampled taxa is provided at the bottom. The dashed lines for *N. tabacum* (Nitab) indicates its allopolyploid nature, with S genome and T genome contributions from *N. sylvestris* (Nisy) and *N. tomentosum* (Nitom), respectively.

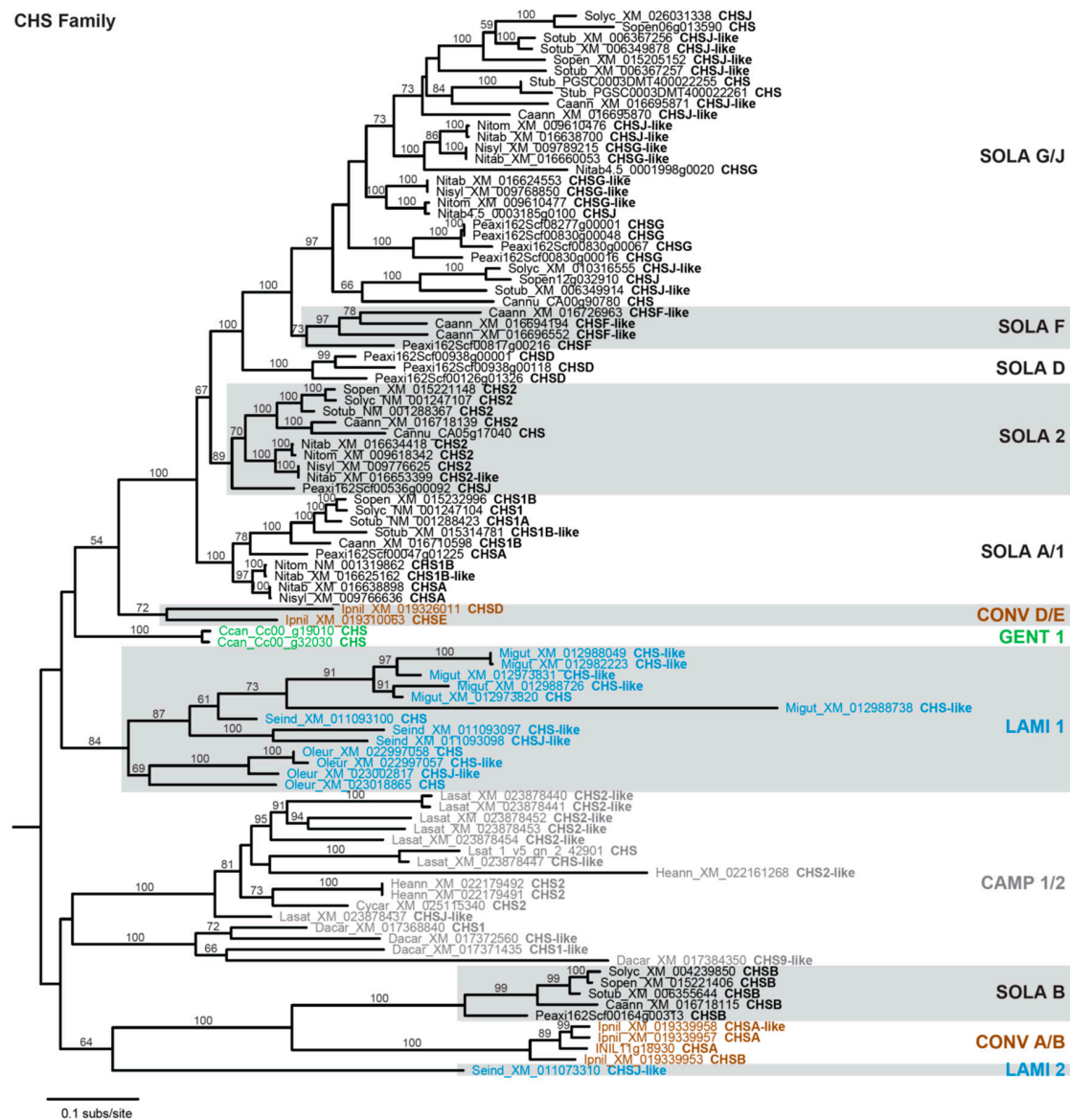
The number of detected homologs varied substantially among FBP enzymes (Figure 1C) and species (Figure 1D). Notably, substantially more homologs were recovered per species for *CHS* compared with other enzymes, due to the fact that *CHS* is a member of a larger chalcone synthase superfamily that includes related proteins with non-*CHS* functions [61,62]. Thus, it is likely that only a fraction of the *CHS* homologs detected in this analysis encode true *CHS* proteins that synthesize chalcones, and hereafter, we will refer to the detected *CHS* homologs as the *CHS* family. Overall, the NCBI and genome-specific pipeline annotations were largely in agreement in terms of the number of homologs per gene and per species, as demonstrated by the majority of homologs having a perfect or imperfect pair between annotation sources, although relatively large discrepancies between annotation results were detected for the *CHS* family, for *F3'5'H*, and for *V. vinifera* (Figure 1C; Figure 1D). These discrepancies were primarily caused by the recent lineage-specific expansions of these two gene families in *V. vinifera* [63,64], with more of these recent duplicates detected in the *V. vinifera* annotation produced by NCBI than in the *V. vinifera* annotation obtained from Phytozome (Figure 1A).

### 3.2. Rampant Lineage-Specific Duplication of Anthocyanin Biosynthesis Genes in Asterids

The results from Figure 1 corroborate that *CHS* has many paralogs in most species and indicate some level of gene duplication for the other FBP enzymes. Phylogenetic analysis was performed to examine the timing of duplication and relationship among these paralogs. Of the seven anthocyanin biosynthesis genes, the *CHS* family exhibits the most dynamic evolutionary history, with multiple paralogous gene groups for most sampled asterids (Figure 2). Six paralogous groups were recovered for Solanaceae, corresponding closely with the A+B+D+F+G+J groups of *CHS* homologs previously defined for *Petunia* [65]. Five of the Solanaceae groups are most closely related to one another, implying a complex history of gene duplications and losses during Solanaceae diversification. The sixth Solanaceae group (SOLA B) contained a single copy for *Capsicum*, *Petunia*, and *Solanum*, indicating a recent loss of this homolog from *Nicotiana*. For *Ipomoea* (the sole representative of Convolvulaceae), two distinct paralogous groups (CONV A/B, CONV D/E) were identified that correspond to the previously described A+B+C and D+E groups [66], and these two clades are sister with the two Solanaceae clades. For Lamiales, one paralog group (LAMI 1) had multiple representatives for all three species, while the second group (LAMI 2) was represented by a single gene from *Sesamum*. Only one paralog group (GENT 1) was obtained for *Coffea*, the lone representative of Gentianales, and for campanulids (CAMP 1/2).

Compared with *CHS*, the remaining FBP genes are less evolutionarily dynamic, with far fewer paralog copies on average, and each tree was largely congruent with organismal phylogeny (Figure 3). For *CHI* (Figure 3A), lineage-specific duplications were detected for *Solanum*, *Petunia*, *Mimulus*, *Helianthus*, and *Cynara*. The *F3H* tree (Figure 3B) revealed lineage-specific duplications for *Capsicum*, *Ipomoea*, *Daucus*, *Mimulus*, and *Olea*. Unexpectedly, campanulids and lamiids were not recovered as monophyletic groups in the *F3H* tree, which suggests either differential loss of two ancestral paralogs or a phylogenetic artifact. The *F3'H* and *F3'5'H* were evaluated together (Figure 3C) as they are members of the CYP75 subfamily of cytochrome P450 genes [67]. Lineage-specific duplications of *F3'H* were apparent for *Mimulus*, *Coffea*, *Helianthus*, *Lactuca*, *N. sylvestris*, and the S genome of *N. tabacum*, while *F3'5'H* homologs were not identified from any asterids except Solanaceae, with duplications affecting *Petunia* and *Capsicum*. For *DFR* (Figure 3D), three *Ipomoea* homologs resulted from recent tandem duplication events [68], and an additional truncated gene was present for *N. sylvestris*, the S genome of *N. tabacum*, and *Olea*. Additionally, a multigene paralogous group was detected for *Daucus* (the *DFR*-related clade). The *Daucus DFR* gene in the species-rich *DFR* clade was orthologous to the previously characterized DcDFR1 [69,70], while the other *DFR*-like paralogs were more similar to DcDFR2, which does not seem to participate in anthocyanin biosynthesis [69]. The ANS tree revealed just a single duplication for *Helianthus* (Figure 3E).

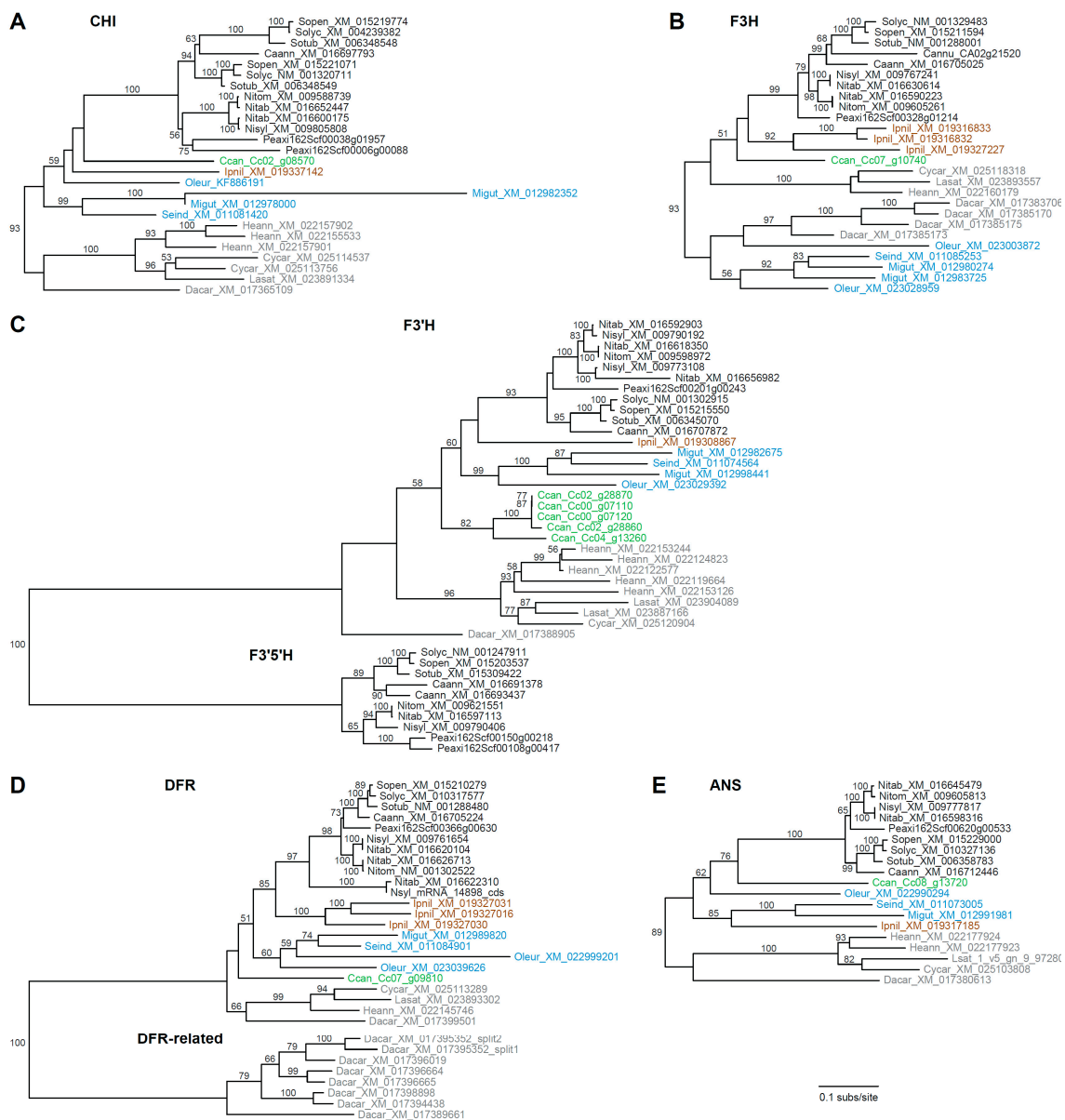




**Figure 2.** Phylogram depicting phylogenetic relationships among *CHS* family homologs. Taxon names include a species abbreviation, a GenBank accession number or gene ID, and the Chalcone synthase (CHS) subfamily designation taken from the accession. Bootstrap values >50% are shown. Monophyletic clades are labeled as described in text and color-coded according to phylogenetic grouping: Campanulids, gray; Convolvulaceae, brown; Gentianales, green; Lamiales, blue; Solanaceae, black. Non-asterisk sequences used to root the tree are not shown.

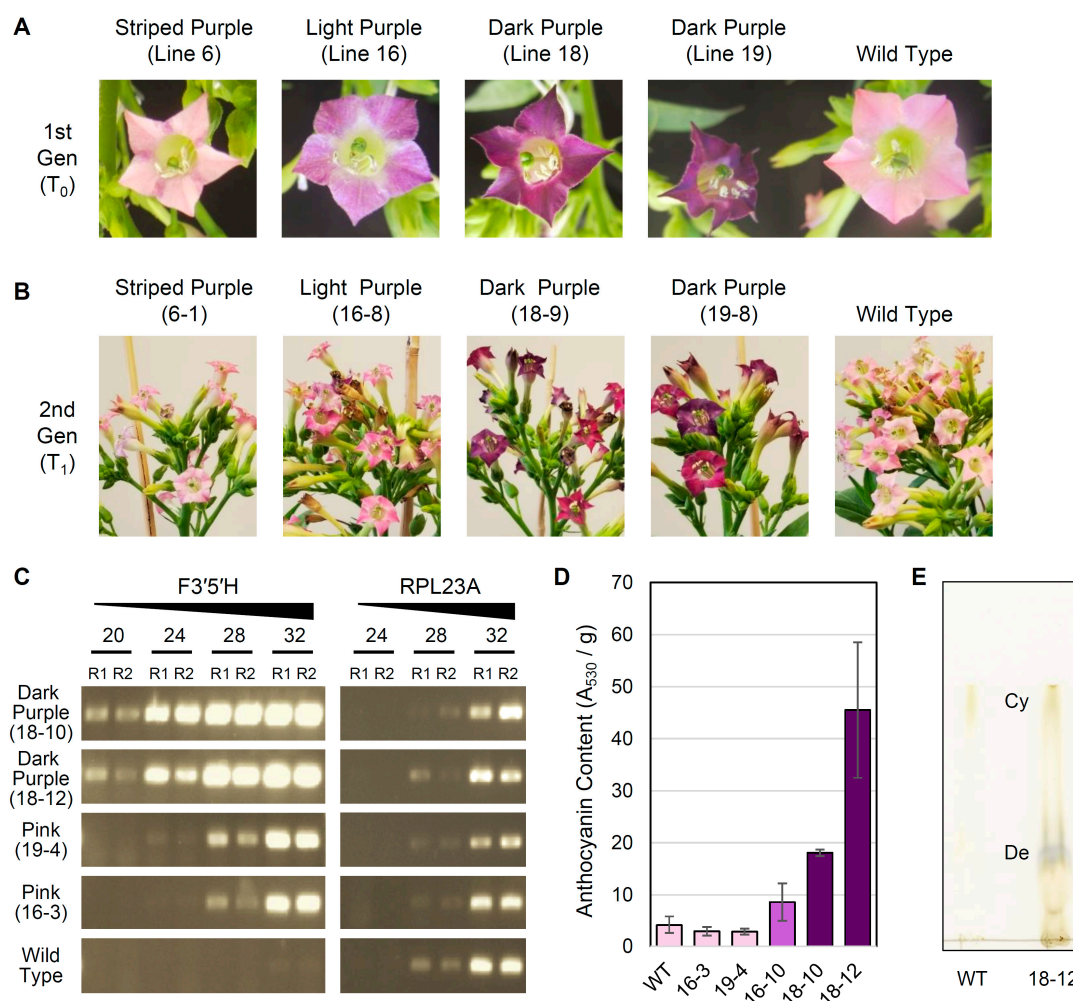
### 3.3. Overexpression of *F3'5'H* Produces Purple Flowers in Tobacco cv. *Xanthi*

The *N. tabacum* genome encodes all seven FBP enzymes for anthocyanin biosynthesis, yet the flowers of most cultivars are pink due to the predominance of cyanidin, as well as its metabolic precursors dihydrokaempferol (DHK) and dihydroquercetin (DHQ), in petals [32]. Because DHK and DHQ are present, it was hypothesized that FBP flux could be altered from cyanidin to delphinidin via introduction of a highly expressed *F3'5'H* gene that produces dihydromyricetin (DHM), the precursor of delphinidin, from DHK and DHQ. Thus, we transformed tobacco cv. *Xanthi* with a construct containing a pansy *F3'5'H* transgene under control of an enhanced 35S promoter. Eighteen regenerated plants ( $T_0$ ) were grown to maturity, with flower colors ranging from WT light pink to dark purple. Expression of *NptII* was confirmed for 12 of 18 plants, of which 7 plants exhibited a solid or striped purple flower phenotype.



**Figure 3.** Phylograms depicting phylogenetic relationships among anthocyanin biosynthesis genes. (A) *CHI*, (B) *F3H*, (C) *F3'H* and *F3'5'H*, (D) *DFR*, and (E) *ANS*. All trees are drawn to the same scale shown at the bottom. Taxon names include a species abbreviation and GenBank accession number or gene ID. Bootstrap values >50% are shown. Major clades are color coded as in Figure 2.

To assess the stability of transformation, seeds from self-pollinated T<sub>0</sub> plants were collected from four lines (Figure 4A) exhibiting a variety of mutant flower color phenotypes, including striped purple (L6), light purple (L16), and dark purple (L18, L19). For each T<sub>0</sub> line, a population of twelve T<sub>1</sub> plants were grown to maturity, and their flower colors were evaluated (Figure 4B; Figure S1). For L18, all T<sub>1</sub> plants had dark purple flowers. For L19, most T<sub>1</sub> plants exhibited either dark or light purple flowers, but three T<sub>1</sub> plants had pink flowers similar to WT. For L16, most T<sub>1</sub> plants had light purple flowers, except that two T<sub>1</sub> plants had pink flowers like WT. For L6, one T<sub>1</sub> plant had dark purple flowers while the remaining T<sub>1</sub> plants either exhibited inconsistent purple striping (present in only some flowers) or pink flowers consistent with WT.



**Figure 4.** Characterization of transgenic tobacco lines. (A) Representative flowers from the T<sub>0</sub> generation and wild type tobacco. (B) Representative inflorescences from the T<sub>1</sub> generation and wild type tobacco. (C) Semi-quantitative reverse transcription PCR (RT-PCR) analysis of transgenic *F3'5'H* and native *RPL23A* expression in T<sub>1</sub> individuals and wild type tobacco. RT-PCR reactions were stopped at 20, 24, 28, and 32 PCR cycles and run on a 1.5% agarose gel, and reactions were performed in replicate (R1, R2). (D) Mean  $\pm$  2 $\times$ SEM for anthocyanin content (measured as absorbance at 530 nm per g tissue) in stage 12 flowers from T<sub>1</sub> individuals and wild type (WT) tobacco. Histogram bars for each sampled individual are color coded according to approximate flower color. (E) Thin-layer chromatography comparing cyanidin (Cy) and delphinidin (De) content in flowers from line 18-12 and WT.

As expected, the relative expression level of the *F3'5'H* transgene was higher in T<sub>1</sub> plants with dark purple flowers than those with pink flowers (Figure 4C), which was indicated by the much earlier appearance of RT-PCR products starting at cycle 20 for the dark purple lines compared with cycle 24 or 28 for the light purple lines, even though the same amount of starting cDNA material was used for each reaction. No major variation was observed in the relative expression level of *RPL23A*, confirming that similar amounts of cDNA were present in each reaction, and no expression of the *F3'5'H* transgene was detected in WT plants. Additionally, relative anthocyanin content (measured by absorbance at 530 nm per g of tissue) correlated well with the intensity of purple coloration in representative T<sub>1</sub> individuals (Figure 4D). Dark purple flowers from two different T<sub>1</sub> plants (18-10 and 18-12) had significantly higher anthocyanin content than WT flowers (unpaired *t*-test,  $P = 0.0001$  and  $P = 0.003$ , respectively). Anthocyanin content was also 2-fold higher in light purple flowers from T<sub>1</sub> line 16-10 compared with WT, although this difference was not significant ( $P = 0.12$ ). For two T<sub>1</sub> lines (16-3 and 19-4) with light pink flowers comparable to WT, floral anthocyanin content was slightly lower

but not significantly different from WT ( $P = 0.13$  and  $P = 0.22$ ). Finally, thin-layer chromatography (Figure 4E) indicated that the dark purple flowers in line 18-12 were the result of a shift in anthocyanin production to primarily delphinidin, with a minor but detectable amount of cyanidin also produced. In contrast, WT tobacco cv. Xanthi contained detectable levels of cyanidin only, as expected [32].

#### 4. Discussion and Conclusions

In this study, we examined the presence and copy number of seven FBP enzymes involved in anthocyanin biosynthesis from 21 solanaceous and related eudicot species. One goal of this study was to compare the consistency of annotations from different sources, such as the NCBI Eukaryotic Genome Annotation Pipeline relative to genome-specific pipeline annotations hosted at Phytozome, the Sol Genomics Network, and genome-specific websites (Table S1). Overall, the large majority of FBP homologs detected from these various annotation sources were in close agreement, but when they differed, homologs from the NCBI annotations were generally longer and more abundant than from the genome-specific databases (Figure 1). These results suggest that reannotation of genome sequences using a unified annotation strategy, such as employed by the NCBI pipeline, may be preferable for improved consistency during comparative genomics research.

A second goal of this research was to assess the diversity of gene content for anthocyanin biosynthesis enzymes among sequenced asterid genomes. In particular, we were interested in assessing whether homology assessment might provide a reliable indicator of gene function. For *CHS*, homology assessment recovered a large gene family exhibiting extensive gene duplication (Figure 2). Neofunctionalization of *CHS* duplicates is widespread [61,71], with particular examples of independent functional transitions from chalcone to non-chalcone production in Asteraceae and *Vitis* [63,72,73]. Subfunctionalization has also been demonstrated for *CHS*; of the multiple *CHS* genes in petunia and *Ipomoea*, only one or two are expressed in flower petals [66,74], while in common bean, several *CHS* homologs are differentially expressed in response to various environmental stressors [75]. Thus, for *CHS*, homology assessment is insufficient to identify the most likely candidates involved in anthocyanin biosynthesis, and molecular characterization is needed to assess protein localization and function. For the other six enzymes, however, phylogenetic analysis of identified homologs produced trees that generally recapitulated organismal relationships, with occasional duplications in particular species (Figure 3). This suggests that many of these genes likely maintain an important role in anthocyanin biosynthesis, with the potential for subfunctionalization in species that contain multiple recent duplicates.

A final goal of this research was to explore the possibility of advancing the Solanaceae FBP as a quantitative model for genotype-to-phenotype interactions of secondary metabolism. One of the most important goals of biological research in the postgenomics era is to establish how genotype determines phenotype. Flower color provides an ideal system to examine this relationship because the genotype (the flavonoid biosynthesis pathway, which produces anthocyanin pigments) is genetically well described, the phenotype (flower color) is directly observable, and the metabolic intermediates (e.g., chalcone, DHK, DHQ, DHM, pelargonidin, cyanidin, and delphinidin) are well known. To be a useful model, simple perturbations should result in predictable phenotypic shifts. In tobacco, the presence of anthocyanin precursors DHK and DHQ suggested that flux could be altered by transgenic overexpression of *F3'5'H*, which converts DHK and DHQ to DHM and commits the pathway to delphinidin biosynthesis. Indeed, several transgenic tobacco lines exhibited clear shifts in flower color from light pink to various purple phenotypes (Figure 4), consistent with previous observations [39,40].

More generally, there are many studies demonstrating flower color evolution resulting from perturbations of the anthocyanin pathway due to functional gene loss [76,77] or to changes in enzyme sequence [78,79] or gene expression levels [80–82]. Several studies have begun to take advantage of the FBP as a quantitative model by evaluating gene expression and metabolites [83–87], yet fine-scale transgenic manipulation of the pathway and quantitative hypothesis testing of such changes on pathway flux are now possible, particularly in Solanaceae where so many genomes and reverse genetic



systems are available. Ultimately, development of the Solanaceae FBP as a manipulable model for secondary metabolism could be used to quantitatively track the effects of complex perturbations on relative expression of enzymes and transcription factors and relative abundance of metabolic intermediates and anthocyanins.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/8/559/s1>, Figure S1: Populations of T1 generation lines in comparison to WT, Table S1: Sources of coding sequence annotations, Table S2: Comparison of anthocyanin biosynthesis genes from NCBI and genome-specific annotations.

**Author Contributions:** Conceptualization, J.P.M.; methodology, Z.L., T.L.V., S.J.S., T.E.C., and J.P.M.; validation, Z.L., T.L.V., M.G.M, S.J.S., and J.P.M.; formal analysis, Z.L., T.L.V., and J.P.M.; visualization, Z.L., T.L.V., and J.P.M., writing—original draft preparation, T.L.V. and J.P.M.; writing—review and editing, Z.L., T.L.V., M.G.M, S.J.S, T.E.C., and J.P.M.

**Funding:** This research was funded in part by the University of Nebraska-Lincoln (to J.P.M.) and the Guangzhou Municipal Science and Technology Program (grant # 201807010092 to Z.L.).

**Acknowledgments:** The authors gratefully acknowledge Pamela Peña and LiLi Hou for advice on plant transformation, Bethany Kiley for assistance with photographing and collecting flowers, Rebecca Roston and Samantha Surber for assistance with chromatography, and Samantha Link and Kandy Hanthorn for help maintaining greenhouse plants.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Grotewold, E. The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* **2006**, *57*, 761–780. [[CrossRef](#)] [[PubMed](#)]
- Feller, A.; Machemer, K.; Braun, E.L.; Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **2011**, *66*, 94–116. [[CrossRef](#)] [[PubMed](#)]
- Xu, W.; Dubos, C.; Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **2015**, *20*, 176–185. [[CrossRef](#)] [[PubMed](#)]
- Falcone Ferreyra, M.L.; Rius, S.P.; Casati, P. Flavonoids: Biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* **2012**, *3*, 222. [[CrossRef](#)] [[PubMed](#)]
- Tanaka, Y.; Sasaki, N.; Ohmiya, A. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant J.* **2008**, *54*, 733–749. [[CrossRef](#)]
- Winkel-Shirley, B. Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **2002**, *5*, 218–223. [[CrossRef](#)]
- Olmstead, R.G.; Bohs, L. A summary of molecular systematic research in Solanaceae: 1982–2006. In Proceedings of the VI International Solanaceae Conference: Genomics Meets Biodiversity, Madison, WI, USA, 23–27 July 2006; Volume 745, pp. 255–268.
- Särkinen, T.; Bohs, L.; Olmstead, R.G.; Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evol. Biol.* **2013**, *13*, 214. [[CrossRef](#)]
- Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S.; Li, R.; The Potato Genome Sequencing Consortium; et al. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189–195.
- Tomato Genome, C. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **2012**, *485*, 635–641. [[CrossRef](#)]
- Bombarely, A.; Rosli, H.G.; Vrebalov, J.; Moffett, P.; Mueller, L.A.; Martin, G.B. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant Microbe Interact.* **2012**, *25*, 1523–1530. [[CrossRef](#)]
- Sierro, N.; Batten, J.N.; Ouadi, S.; Bovet, L.; Goepfert, S.; Bakaher, N.; Peitsch, M.C.; Ivanov, N.V. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* **2013**, *14*, R60. [[CrossRef](#)] [[PubMed](#)]
- Bolger, A.; Scossa, F.; Bolger, M.E.; Lanz, C.; Maumus, F.; Tohge, T.; Quesneville, H.; Alseekh, S.; Sorensen, I.; Lichtenstein, G.; et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **2014**, *46*, 1034–1038. [[CrossRef](#)] [[PubMed](#)]

14. Hirakawa, H.; Shirasawa, K.; Miyatake, K.; Nunome, T.; Negoro, S.; Ohyama, A.; Yamaguchi, H.; Sato, S.; Isobe, S.; Tabata, S.; et al. Draft genome sequence of eggplant (*Solanum melongena* L.): The representative solanum species indigenous to the old world. *DNA Res.* **2014**, *21*, 649–660. [[CrossRef](#)] [[PubMed](#)]
15. Kim, S.; Park, M.; Yeom, S.I.; Kim, Y.M.; Lee, J.M.; Lee, H.A.; Seo, E.; Choi, J.; Cheong, K.; Kim, K.T.; et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **2014**, *46*, 270–278. [[CrossRef](#)] [[PubMed](#)]
16. Qin, C.; Yu, C.; Shen, Y.; Fang, X.; Chen, L.; Min, J.; Cheng, J.; Zhao, S.; Xu, M.; Luo, Y.; et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5135–5140. [[CrossRef](#)] [[PubMed](#)]
17. Sierro, N.; Batten, J.N.; Ouadi, S.; Bakaher, N.; Bovet, L.; Willig, A.; Goepfert, S.; Peitsch, M.C.; Ivanov, N.V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **2014**, *5*, 3833. [[CrossRef](#)] [[PubMed](#)]
18. Bombarely, A.; Moser, M.; Amrad, A.; Bao, M.; Bapaume, L.; Barry, C.S.; Bliet, M.; Boersma, M.R.; Borghi, L.; Bruggmann, R.; et al. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* **2016**, *2*, 16074. [[CrossRef](#)]
19. Edwards, K.D.; Fernandez-Pozo, N.; Drake-Stowe, K.; Humphry, M.; Evans, A.D.; Bombarely, A.; Allen, F.; Hurst, R.; White, B.; Kernodle, S.P.; et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* **2017**, *18*, 448. [[CrossRef](#)]
20. Xu, S.; Brockmoller, T.; Navarro-Quezada, A.; Kuhl, H.; Gase, K.; Ling, Z.; Zhou, W.; Kreitzer, C.; Stanke, M.; Tang, H.; et al. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6133–6138. [[CrossRef](#)]
21. Hulse-Kemp, A.M.; Maheshwari, S.; Stoffel, K.; Hill, T.A.; Jaffe, D.; Williams, S.R.; Weisenfeld, N.; Ramakrishnan, S.; Kumar, V.; Shah, P.; et al. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* **2018**, *5*, 4. [[CrossRef](#)]
22. Hoshino, A.; Jayakumar, V.; Nitasaka, E.; Toyoda, A.; Noguchi, H.; Itoh, T.; Shin, I.T.; Minakuchi, Y.; Koda, Y.; Nagano, A.J.; et al. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat. Commun.* **2016**, *7*, 13295. [[CrossRef](#)] [[PubMed](#)]
23. Hellsten, U.; Wright, K.M.; Jenkins, J.; Shu, S.; Yuan, Y.; Wessler, S.R.; Schmutz, J.; Willis, J.H.; Rokhsar, D.S. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19478–19482. [[CrossRef](#)] [[PubMed](#)]
24. Wang, L.; Han, X.; Zhang, Y.; Li, D.; Wei, X.; Ding, X.; Zhang, X. Deep resequencing reveals allelic variation in *Sesamum indicum*. *BMC Plant Biol.* **2014**, *14*, 225. [[CrossRef](#)] [[PubMed](#)]
25. Denoeud, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C.; Alberti, A.; Anthony, F.; Aprea, G.; et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014**, *345*, 1181–1184. [[CrossRef](#)] [[PubMed](#)]
26. Cruz, F.; Julca, I.; Gomez-Garrido, J.; Loska, D.; Marcet-Houben, M.; Cano, E.; Galan, B.; Frias, L.; Ribeca, P.; Derdak, S.; et al. Genome sequence of the olive tree, *Olea europaea*. *Gigascience* **2016**, *5*, 29. [[CrossRef](#)] [[PubMed](#)]
27. Unver, T.; Wu, Z.; Sterck, L.; Turktas, M.; Lohaus, R.; Li, Z.; Yang, M.; He, L.; Deng, T.; Escalante, F.J.; et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9413–E9422. [[CrossRef](#)] [[PubMed](#)]
28. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.; Iovene, M.; Sanseverino, W.; Cavagnaro, P.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657–666. [[CrossRef](#)]
29. Badouin, H.; Gouzy, J.; Grassa, C.J.; Murat, F.; Staton, S.E.; Cottret, L.; Lelandais-Briere, C.; Owens, G.L.; Carrere, S.; Mayjonade, B.; et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **2017**, *546*, 148–152. [[CrossRef](#)]
30. Reyes-Chin-Wo, S.; Wang, Z.; Yang, X.; Kozik, A.; Arikait, S.; Song, C.; Xia, L.; Froenicke, L.; Lavelle, D.O.; Truco, M.J.; et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* **2017**, *8*, 14953. [[CrossRef](#)]
31. Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; et al. The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. *Sci. Rep.* **2016**, *6*, 19427. [[CrossRef](#)]



32. McCarthy, E.W.; Berardi, A.E.; Smith, S.D.; Litt, A. Related allopolyploids display distinct floral pigment profiles and transgressive pigments. *Am. J. Bot.* **2017**, *104*, 92–101. [[CrossRef](#)] [[PubMed](#)]
33. Azuma, K.; Ohyama, A.; Ippoushi, K.; Ichianagi, T.; Takeuchi, A.; Saito, T.; Fukuoka, H. Structures and antioxidant activity of anthocyanins in many accessions of eggplant and its related species. *J. Agric. Food Chem.* **2008**, *56*, 10154–10159. [[CrossRef](#)] [[PubMed](#)]
34. Gates, D.J.; Olson, B.; Clemente, T.E.; Smith, S.D. A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *Ipomoea*. *New Phytol.* **2018**, *217*, 1346–1356. [[CrossRef](#)]
35. Rodriguez-Saona, L.E.; Giusti, M.M.; Wrolstad, R.E. Anthocyanin pigment composition of red-fleshed potatoes. *J. Food Sci.* **1998**, *63*, 458–465. [[CrossRef](#)]
36. Lewis, C.E.; Walker, J.R.; Lancaster, J.E.; Sutton, K.H. Determination of anthocyanins, flavonoids and phenolic acids in potatoes. I: Coloured cultivars of *Solanum tuberosum* L. *J. Sci. Food Agric.* **1998**, *77*, 45–57. [[CrossRef](#)]
37. Liu, Y.; Tikunov, Y.; Schouten, R.E.; Marcelis, L.F.M.; Visser, R.G.F.; Bovy, A. Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: A review. *Front. Chem.* **2018**, *6*, 52. [[CrossRef](#)] [[PubMed](#)]
38. Meyer, P.; Heidmann, I.; Forkmann, G.; Saedler, H. A new petunia flower colour generated by transformation of a mutant with a maize gene. *Nature* **1987**, *330*, 677–678. [[CrossRef](#)] [[PubMed](#)]
39. Shimada, Y.; Nakano-Shimada, R.; Ohbayashi, M.; Okinaka, Y.; Kiyokawa, S.; Kikuchi, Y. Expression of chimeric P450 genes encoding flavonoid-3', 5'-hydroxylase in transgenic tobacco and petunia plants. *FEBS Lett.* **1999**, *461*, 241–245. [[CrossRef](#)]
40. Okinaka, Y.; Shimada, Y.; Nakano-Shimada, R.; Ohbayashi, M.; Kiyokawa, S.; Kikuchi, Y. Selective accumulation of delphinidin derivatives in tobacco using a putative flavonoid 3',5'-hydroxylase cDNA from *Campanula medium*. *Biosci. Biotechnol. Biochem.* **2003**, *67*, 161–165. [[CrossRef](#)]
41. Nakatsuka, T.; Nishihara, M.; Mishiba, K.; Yamamura, S. Heterologous expression of two gentian cytochrome P450 genes can modulate the intensity of flower pigmentation in transgenic tobacco plants. *Mol. Breed.* **2006**, *17*, 91–99. [[CrossRef](#)]
42. Vanderkrol, A.R.; Lenting, P.E.; Veenstra, J.; Vandermeer, I.M.; Koes, R.E.; Gerats, A.G.M.; Mol, J.N.M.; Stuitje, A.R. An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. *Nature* **1988**, *333*, 866–869. [[CrossRef](#)]
43. Nishihara, M.; Nakatsuka, T.; Yamamura, S. Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. *FEBS Lett.* **2005**, *579*, 6074–6078. [[CrossRef](#)] [[PubMed](#)]
44. Nakatsuka, T.; Pitaksutheepong, C.; Yamamura, S.; Nishihara, M. Induction of differential flower pigmentation patterns by RNAi using promoters with distinct tissue-specific activity. *Plant Biotechnol. Rep.* **2007**, *1*, 251–257. [[CrossRef](#)]
45. Holton, T.A.; Brugliera, F.; Tanaka, Y. Cloning and expression of flavonol synthase from *Petunia hybrida*. *Plant J.* **1993**, *4*, 1003–1010. [[CrossRef](#)] [[PubMed](#)]
46. Fischer, R.; Budde, I.; Hain, R. Stilbene synthase gene expression causes changes in flower colour and male sterility in tobacco. *Plant J.* **1997**, *11*, 489–498. [[CrossRef](#)]
47. Davies, K.M.; Bloor, S.J.; Spiller, G.B.; Deroles, S.C. Production of yellow colour in flowers: Redirection of flavonoid biosynthesis in *Petunia*. *Plant J.* **1998**, *13*, 259–266. [[CrossRef](#)]
48. Davies, K.M.; Schwinn, K.E.; Deroles, S.C.; Manson, D.G.; Lewis, D.H.; Bloor, S.J.; Bradley, J.M. Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. *Euphytica* **2003**, *131*, 259–268. [[CrossRef](#)]
49. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)]
50. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
51. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [[CrossRef](#)]
52. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]

53. Katsumoto, Y.; Fukuchi-Mizutani, M.; Fukui, Y.; Brugliera, F.; Holton, T.A.; Karan, M.; Nakamura, N.; Yonekura-Sakakibara, K.; Togami, J.; Pigeaire, A.; et al. Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. *Plant Cell Physiol.* **2007**, *48*, 1589–1600. [[CrossRef](#)] [[PubMed](#)]
54. Tanaka, Y.; Brugliera, F. Flower colour and cytochromes P450. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2013**, *368*, 20120432. [[CrossRef](#)] [[PubMed](#)]
55. Hajdukiewicz, P.; Svab, Z.; Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **1994**, *25*, 989–994. [[CrossRef](#)] [[PubMed](#)]
56. Koncz, C.; Schell, J. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **1986**, *204*, 383–396. [[CrossRef](#)]
57. Clemente, T. Nicotiana (*Nicotiana tobaccum*, *Nicotiana benthamiana*). *Methods Mol. Biol.* **2006**, *343*, 143–154. [[PubMed](#)]
58. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
59. Sigmon, B.A.; Adams, R.P.; Mower, J.P. Complete chloroplast genome sequencing of vetiver grass (*Chrysopogon zizanioides*) identifies markers that distinguish the non-fertile ‘Sunshine’ cultivar from other accessions. *Ind. Crops Prod.* **2017**, *108*, 629–635. [[CrossRef](#)]
60. Hepburn, N.J.; Schmidt, D.W.; Mower, J.P. Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol. Biol. Evol.* **2012**, *29*, 3111–3120. [[CrossRef](#)]
61. Jiang, C.; Kim, S.Y.; Suh, D.Y. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylogenet. Evol.* **2008**, *49*, 691–701. [[CrossRef](#)]
62. Abe, I.; Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **2010**, *27*, 809–838. [[CrossRef](#)]
63. Parage, C.; Tavares, R.; Rety, S.; Baltenweck-Guyot, R.; Poutaraud, A.; Renault, L.; Heintz, D.; Lugan, R.; Marais, G.A.; Aubourg, S.; et al. Structural, functional, and evolutionary analysis of the unusually large stilbene synthase gene family in grapevine. *Plant Physiol.* **2012**, *160*, 1407–1419. [[CrossRef](#)] [[PubMed](#)]
64. Falginella, L.; Castellarin, S.D.; Testolin, R.; Gambetta, G.A.; Morgante, M.; Di Gaspero, G. Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. *BMC Genom.* **2010**, *11*, 562. [[CrossRef](#)] [[PubMed](#)]
65. Koes, R.E.; Spelt, C.E.; Mol, J.N.M. The chalcone synthase multigene family of *Petunia hybrida* (V30)-differential, light-regulated expression during flower development and uv-light induction. *Plant Mol. Biol.* **1989**, *12*, 213–225. [[CrossRef](#)] [[PubMed](#)]
66. Durbin, M.L.; McCaig, B.; Clegg, M.T. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* **2000**, *42*, 79–92. [[CrossRef](#)] [[PubMed](#)]
67. Ayabe, S.-I.; Akashi, T. Cytochrome P450s in flavonoid metabolism. *Phytochem. Rev.* **2006**, *5*, 271–282. [[CrossRef](#)]
68. Inagaki, Y.; Johzuka-Hisatomi, Y.; Mori, T.; Takahashi, S.; Hayakawa, Y.; Peyachoknagul, S.; Ozeki, Y.; Iida, S. Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and common morning glories. *Gene* **1999**, *226*, 181–188. [[CrossRef](#)]
69. Hirner, A.A.; Veit, S.; Seitz, H.U. Regulation of anthocyanin biosynthesis in UV-A-irradiated cell cultures of carrot and in organs of intact carrot plants. *Plant Sci.* **2001**, *161*, 315–322. [[CrossRef](#)]
70. Xu, Z.S.; Huang, Y.; Wang, F.; Song, X.; Wang, G.L.; Xiong, A.S. Transcript profiling of structural genes involved in cyanidin-based anthocyanin biosynthesis between purple and non-purple carrot (*Daucus carota* L.) cultivars reveals distinct patterns. *BMC Plant Biol.* **2014**, *14*, 262. [[CrossRef](#)]
71. Tropf, S.; Lanz, T.; Rensing, S.A.; Schroder, J.; Schroder, G. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *J. Mol. Evol.* **1994**, *38*, 610–618. [[CrossRef](#)]
72. Helariutta, Y.; Kotilainen, M.; Elomaa, P.; Kalkkinen, N.; Bremer, K.; Teeri, T.H.; Albert, V.A. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: Evolution with substrate change and catalytic simplification. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9033–9038. [[CrossRef](#)] [[PubMed](#)]
73. Yang, J.; Huang, J.; Gu, H.; Zhong, Y.; Yang, Z. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). *Mol. Biol. Evol.* **2002**, *19*, 1752–1759. [[CrossRef](#)]

74. Koes, R.E.; Spelt, C.E.; Reif, H.J.; van den Elzen, P.J.; Veltkamp, E.; Mol, J.N. Floral tissue of *Petunia hybrida* (V30) expresses only one member of the chalcone synthase multigene family. *Nucleic Acids Res.* **1986**, *14*, 5229–5239. [[CrossRef](#)] [[PubMed](#)]
75. Ryder, T.B.; Hedrick, S.A.; Bell, J.N.; Liang, X.W.; Clouse, S.D.; Lamb, C.J. Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* **1987**, *210*, 219–233. [[CrossRef](#)] [[PubMed](#)]
76. Smith, S.D.; Rausher, M.D. Gene loss and parallel evolution contribute to species difference in flower color. *Mol. Biol. Evol.* **2011**, *28*, 2799–2810. [[CrossRef](#)] [[PubMed](#)]
77. Ishiguro, K.; Taniguchi, M.; Tanaka, Y. Functional analysis of *Antirrhinum kelloggii* flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes; critical role in flower color and evolution in the genus *Antirrhinum*. *J. Plant Res.* **2012**, *125*, 451–456. [[CrossRef](#)] [[PubMed](#)]
78. Smith, S.D.; Wang, S.; Rausher, M.D. Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Mol. Biol. Evol.* **2013**, *30*, 602–612. [[CrossRef](#)] [[PubMed](#)]
79. Wu, C.A.; Streisfeld, M.A.; Nutter, L.I.; Cross, K.A. The genetic basis of a rare flower color polymorphism in *Mimulus lewisii* provides insight into the repeatability of evolution. *PLoS ONE* **2013**, *8*, e81173. [[CrossRef](#)] [[PubMed](#)]
80. Heppel, S.C.; Jaffe, F.W.; Takos, A.M.; Schellmann, S.; Rausch, T.; Walker, A.R.; Bogs, J. Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Mol. Biol.* **2013**, *82*, 457–471. [[CrossRef](#)] [[PubMed](#)]
81. Liu, X.J.; Chuang, Y.N.; Chiou, C.Y.; Chin, D.C.; Shen, F.Q.; Yeh, K.W. Methylation effect on chalcone synthase gene expression determines anthocyanin pigmentation in floral tissues of two *Oncidium* orchid cultivars. *Planta* **2012**, *236*, 401–409. [[CrossRef](#)] [[PubMed](#)]
82. Butelli, E.; Titta, L.; Giorgio, M.; Mock, H.P.; Matros, A.; Peterek, S.; Schijlen, E.G.; Hall, R.D.; Bovy, A.G.; Luo, J.; et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **2008**, *26*, 1301–1308. [[CrossRef](#)] [[PubMed](#)]
83. Jin, X.; Huang, H.; Wang, L.; Sun, Y.; Dai, S. Transcriptomics and metabolite analysis reveals the molecular mechanism of anthocyanin biosynthesis branch pathway in different *Senecio cruentus* cultivars. *Front. Plant Sci.* **2016**, *7*, 1307. [[CrossRef](#)] [[PubMed](#)]
84. Zorenc, Z.; Veberic, R.; Slatnar, A.; Koron, D.; Miosic, S.; Chen, M.H.; Haselmair-Gosch, C.; Halbwirth, H.; Mikulic-Petkovsek, M. A wild 'albino' bilberry (*Vaccinium myrtillus* L.) from Slovenia shows three bottlenecks in the anthocyanin pathway and significant differences in the expression of several regulatory genes compared to the common blue berry type. *PLoS ONE* **2017**, *12*, e0190246. [[CrossRef](#)] [[PubMed](#)]
85. Zhang, C.; Li, X.; Zhan, Z.; Cao, L.; Zeng, A.; Chang, G.; Liang, Y. Transcriptome sequencing and metabolism analysis reveals the role of cyanidin metabolism in dark-red onion (*Allium cepa* L.) bulbs. *Sci. Rep.* **2018**, *8*, 14109. [[CrossRef](#)] [[PubMed](#)]
86. Deguchi, A.; Ohno, S.; Hosokawa, M.; Tatsuzawa, F.; Doi, M. Endogenous post-transcriptional gene silencing of flavone synthase resulting in high accumulation of anthocyanins in black dahlia cultivars. *Planta* **2013**, *237*, 1325–1335. [[CrossRef](#)] [[PubMed](#)]
87. Lou, Q.; Liu, Y.; Qi, Y.; Jiao, S.; Tian, F.; Jiang, L.; Wang, Y. Transcriptome sequencing and metabolite analysis reveals the role of delphinidin metabolism in flower colour in grape hyacinth. *J. Exp. Bot.* **2014**, *65*, 3157–3164. [[CrossRef](#)]

