

# Co-Expression Network Analyses of Anthocyanin Biosynthesis Genes in *Ruellia* (Wild Petunias; Acanthaceae)

Erin Tripp (✉ [erin.tripp@colorado.edu](mailto:erin.tripp@colorado.edu))

University of Colorado Boulder <https://orcid.org/0000-0001-9340-8723>

Yongbin Zhuang

Shandong University

---

## Research article

**Keywords:** anthocyanin, evolution flower color, hybrid transcriptome-RADseq phylogeny, MYBs, shift, transition

**Posted Date:** August 20th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-58072/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background.** Anthocyanins are major pigments contributing to flower coloration and as such knowledge of molecular architecture underlying the anthocyanin biosynthetic pathway (ABP) is key to understanding flower color diversification. To identify ABP structural genes and associated regulatory networks, we sequenced 16 transcriptomes generated from 10 species of *Ruellia* and then conducted co-expression analyses among resulting data.

**Results.** Complete coding sequences for 12 candidate structural loci representing eight genes plus nine candidate regulatory loci were assembled. Analysis of non-synonymous/synonymous (dn/ds) mutation rates indicated all identified loci are under purifying selection, suggesting overall selection to prevent the accumulation of deleterious mutations. Additionally, upstream enzymes have lower rates of molecular evolution compared to downstream enzymes. However, site-specific tests of selection yielded evidence for positive selection at several sites, including four in *F3'H2* and five in *DFR3*, and these sites are located in protein binding regions. A species-level phylogenetic tree constructed using a newly implemented hybrid transcriptome–RADseq approach implicates numerous flower color transitions among the 10 species. We found evidence of both regulatory and structural mutations to *F3'5'H* in helping to explain the evolution of red flowers from purple-flowered ancestors.

**Conclusions.** Sequence comparisons and co-expression analyses of ABP loci revealed that mutations in regulatory loci are likely to play a greater role in flower color transitions in *Ruellia* compared to mutations in underlying structural genes.

## Background

Across angiosperms, flower color is determined by a number of different factors including the presence of one or more pigment classes (e.g., anthocyanins, non-anthocyanin flavonoids, carotenoids including xanthophylls, betalains, and chlorophyll) and environmental factors including pH and UV exposure (Zhao et al., 2011; Zhao and Tao, 2015). Among these different classes, anthocyanins are widely appreciated as one of the most important and evolutionary widespread plant pigment pathways (Weiss, 1995; Petroni and Tonelli, 2001; Tanaka et al., 2008; Heppel et al., 2013). Anthocyanins are water-soluble pigments that belong to the flavonoid biosynthetic pathway. The three primary aglycone forms of anthocyanin plus three derivatives—pelargonidin, cyanidin (including peonidin), and delphinidin (including petunidin and malvidin)—are together responsible for red, pink, and purple to blue pigmentation, respectively (Zhang et al., 2014).

The Anthocyanin Biosynthetic Pathway (ABP) has been extensively studied from a molecular perspective and is highly conserved across flowering plants. Studies in model species such as petunia, snapdragon, and *Arabidopsis* have identified eight structural genes encoding the following core ABP enzymes: chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonoid 3'5'-hydroxylase (F3'5'H), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-3-O-glucosyltransferases (UF3GT). The temporal and spatial expression of these eight genes as well as any structural variations due to mutations in these genes plus differential binding of pathway intermediates to enzymes depending on substrate specificity yields flower color variation (Tanaka et al., 2008; Heppel et al., 2013).

Anthocyanin biosynthesis is regulated primarily at the transcriptional level by a complex of three classes of transcription factors—MYB, basic helix-loop-helix (bHLH), and WD-40 repeat (WDR) proteins—collected referred to as the MBW complex (Patzlaff et al., 2003). Compared to bHLH and WDR, MYBs play a central role in recognizing specific target genes (Baudry et al., 2004; Nesi et al., 2001; Zhao et al., 2013). A plethora of studies have identified MYB homologs and their functions in various plant families. An R2R3-MYB encoding gene named *VvmybA1* is a key regulator of *UFGT* in grape skin and other parts of the grape plant (Yakushiji et al., 2006). In *Antirrhinum majus*, expression of *F3H*, *DFR*, *ANS* and *UF3GT* in flowers is regulated by three closely related R2R3-MYBs named *Rosea1*, *Rosea2* and *Venosa* (Goodrich et al., 1992; Schwinn et al., 2006). Three MYBs isolated from apple, *MdMYB1*, *MdMYB10* and *MdMYB*, and 'MYB gene *Ruby*' from *citrus*, facilitate the accumulation of anthocyanins in the fruit (Espley et al., 2007; Butelli et al., 2012). MYBs furthermore play tissue- and location-specific roles in flower coloration. In *Mimulus lewisii*, two R2R3-MYBs named *PELAN* and *NEGAN* regulate anthocyanin biosynthesis in petal lobes and nectar guide spots, respectively (Yuan et al., 2014).

Significant progress made in elucidating ABP regulatory mechanisms in non-model plants has been possible owing to the conserved nature of ABP enzymes and their MBW regulators. However, study of ABP regulation through homolog-based approaches limits potential identification of novel, regulatory elements because most identified MYBs are highly conserved and belong to the same clade in phylogenetic analyses. Somatic mutations of identified MYBs generally affect the expression of one or a few key structural genes (Chiu et al., 2010), suggesting the existence of other unidentified regulators. In the present study, in addition to identification of genetic elements involved in ABP expression through a homolog-based search method, we undertake co-expression network association analyses (Figure S1) using 16 newly generated transcriptomes from 10 closely related species of *Ruellia* (Wild Petunias: Acanthaceae) to (1) yield a first, comprehensive estimate of candidate structural and regulatory loci underlying flower coloration in *Ruellia*, (2) examine molecular evolution of these loci, (3) assess whether mutations to structural vs. regulatory elements have likely been more important during flower color evolution in the group, (4) identify blocks of the ABP likely under co-regulation, and (5) identify potentially novel candidate transcription factors involved in the ABP.

## Results

### HPLC analysis of anthocyanin accumulation in leaf and petal tissues

To phenotype ABP pigments that underlie flower color variation in *Ruellia*, 10 species with distinct petal colors (four purple, three red, and three yellow) were selected based on visual inspection coupled followed by floral color delimitation into one of several discrete categories. The latter was accomplished via floral reflectance measured using an Ocean Optics JAZ-COMBO Spectrometer (Fig. 1A; Data S1). Presence/absence of cyanidin, delphinidin, pelargonidin and their derivatives were measured in both petal and leaf tissues using HPLC (Data S2). No anthocyanins were detected in petal tissues of the three yellow-flowered species (*Ruellia bourgaei*, *R. speciosa*, and *R. lutea*). Two of the three red-flowered species (*R. brevifolia*, *R. fulgida*) produced only pelargonidins whereas in the third red-flowered species (*R. elegans*), both pelargonidin and cyanidin were detected. Two of the purple-flowered species (*R. breedlovei*, *R. hirsuto-glandulosa*) produced only delphinidins whereas the remaining two purple-flowered species (*R. simplex*, *R. longepetiolata*) manufactured both delphinidins and cyanidins (Fig. 1A, Data S2). No species produced the combination of pelargonidin and delphinidin pigments. In leaf tissues, cyanidins were found in all species except *R. bourgaei* and *R. hirsuto-glandulosa* (Data S2).

### De novo assembly of tissue-specific *Ruellia* transcriptomes

To help elucidate molecular mechanisms regulating anthocyanin accumulation, we analyzed data from 12 newly sequenced transcriptome libraries (eight from flower petals, four from leaf tissues) together with data from four previously published libraries (two from flower petals, two from leaf tissues; Zhuang and Tripp, 2017). Quality statistics for tissue-specific assemblies using the best TransRate score are shown in Table 1. On average,  $165,860 \pm 28,003$  transcripts were assembled with an average N50 of  $1,611 \pm 194$  nt, and similar GC contents were found across all libraries ( $42.50 \pm 0.53\%$ ). TransDecoder identified  $73614 \pm 18417$  Open Reading Frames (ORFs) in assembled transcriptomes. Representativeness of each assembly was evaluated by mapping error-corrected reads back to the assembly. A high mapping rate ( $95.96 \pm 2.10\%$ ) was obtained across all assemblies, and among mapped reads, nearly all ( $98.78 \pm 1.04\%$ ) could be mapped concordantly as read pairs. We also evaluated transcriptome completeness by searching for single-copy orthologs of highly conserved plant genes in each assembly using BUSCO v2 (Simão et al., 2015). On average, assembled transcriptomes, regardless whether derived from petal or leaf tissue, contained a high percentage ( $74.06 \pm 5.91\%$ ) of the 1,440 groups of conserved plant genes deposited in the BUSCO database (Table 1); this high percentage further reflects the completeness of our transcriptome assemblies.

Table 1  
Summary statistics of assembled *Ruellia* transcriptomes.

Species	Tissue	# of transcripts	N50 (nt)	GC (%)	# of ORF	BUSCO (%)	Mapping rate (%)
<i>R. bourgaei</i>	petal	143979	1613	43.41	84262	82.60	97.69
<i>R. lutea</i>	petal	202066	1782	42.11	72821	76.67	96.78
<i>R. speciosa</i>	petal	183957	1353	42.66	53524	68.67	90.99
<i>R. fulgida</i>	petal	202811	1911	41.85	82087	78.40	96.72
<i>R. elegans</i>	petal	155972	1313	42.71	57097	65.76	93.43
<i>R. brevifolia</i>	petal	171861	1909	42.68	99328	76.18	98.06
<i>R. breedlovei</i>	petal	162669	1418	42.61	54982	63.82	92.84
<i>R. hirsuto-glandulosa</i>	petal	174679	1690	42.10	64848	73.40	94.57
<i>R. longepetiolata</i>	petal	103568	1718	43.27	62315	62.56	97.08
<i>R. simplex</i>	petal	187555	1769	42.67	107402	69.44	97.27
<i>R. lutea</i>	leaf	149699	1741	41.99	70772	76.25	98.03
<i>R. speciosa</i>	leaf	141301	1308	42.06	54952	75.48	94.47
<i>R. fulgida</i>	leaf	125607	1551	42.62	57570	75.00	97.84
<i>R. elegans</i>	leaf	195570	1522	41.68	76339	73.76	95.91
<i>R. hirsuto-glandulosa</i>	leaf	168329	1626	42.17	70680	76.25	96.63
<i>R. simplex</i>	leaf	184144	1563	43.35	108853	80.69	97.05
Mean		165860	1611	42.50	73614	74.06	95.96
SD		28003	194	0.53	18417	5.91	2.10

## Reconstruction of phylogenetic relationships in *Ruellia*

To resolve phylogenetic relationships among our 10 species of *Ruellia*, we employed a new approach that involves retrieving RAD loci from transcriptome data. Phylogenies constructed using assembled transcriptome data sometimes face challenges due to complications associated with pseudogenes and paralogous loci, resulting in potentially misassembled contigs (Wen et al., 2013). We thus implemented a hybrid transcriptome–RADseq approach by phylogenetic tree building that made use of a subset of error-corrected reads having the *EcoRI* GAATTC overhang. These reads were treated in a manner similar to RADtags and were analyzed using pyRAD (Eaton, 2014). The resulting average read length was 78nt, which is suitable for phylogenetic analysis as it has been shown that large phylogenetic tree can be built accurately from even short (50 bp average) sequences (Hittinger et al., 2010). A total of 110,263 loci were concatenated and used for phylogenetic inference via maximum likelihood implemented in RAxML v8 (Stamatakis, 2014). Branch support as assessed via 100 bootstrap replicates was  $\geq 75\%$  for all nodes (most nodes with  $\geq 95\%$  support), and the resulting phylogeny composed of two major clades (Fig. 1A) mirrors relationships recovered in prior phylogenetic studies in the genus (Tripp and Tsai 2017). We specifically did not conduct ancestral state reconstructions of flower color and estimations of color transitions due to very few species terminals (e.g., Salisbury and Kim 2001), especially in a taxonomically species-rich genus such as *Ruellia* (ca. 350 species worldwide). However, the 10-tip phylogeny suggests four potential color transitions: in Clade I, one from purple to red and one from red to yellow flowers, and in Clade II, one from red to purple and one from purple to yellow flowers.

## PCA analysis of leaf and petal transcriptomes

To assess overall expression level similarities among transcriptomes, we conducted a principal component analysis (PCA) on normalized gene count data based on combined petal and leaf transcriptome data from *R. simplex*, which had the best assembly statistics among all species (Table 1). After removing transcripts with low expression levels (see methods), a total of 60,626 transcripts were used in PCA analysis. Results demonstrate that flower petal vs. leaf transcripts were separated into two clusters (Fig. 1B). Additionally, two subclusters among the flower petal transcripts can be identified and correspond to two clades resolved in our phylogeny (Fig. 1A): Clade I (*R. elegans*, *R. speciosa*, *R. hisuto-glandulosa*, *R. bourgaei*) and Clade II (*R. lutea*, *R. fulgida*, *R. breedlovei*, *R. brevifolia*).

## Identification and co-expression analysis of candidate ABP structural genes

Our BLAST + search of Trinity-assembled *Ruellia* transcriptomes against known ABP genes identified 12 orthologous candidate ABP structural genes in *Ruellia*. Among these, one copy each of *CHS*, *CHI*, *F3H*, *F3'5'H*, *ANS* and *UF3GT* was identified whereas three copies each of *F3'H* and *DFR* were recovered. Subsequent reference-based gene assembly using SPAdes v3.10.1 (Bankevich et al., 2012) recovered full-length cDNA of ABP genes in most of the species (Figure S2 and Table S1). We then measured expression levels of the 12 candidate genes in FPKM (fragments per kilobase of transcript per million fragments sequenced) and compared results in a tissue-specific manner. Results indicate that expression of *DFR1* and *F3'5'H* was restricted to petal tissues whereas expression of all other genes occurred in one or more leaf samples (Fig. 2).

We further characterized the multiple copies of *F3'H* and *DFR* identified among our *Ruellia* transcriptomes by comparing translated amino acid sequence across samples (Figure S2) as well as expression patterns measured at five flower developmental stages in *Ruellia simplex* using transcript-specific primers and qPCR (Fig. 3, Table S2). Two of the three *F3'H* copies (*F3'H1*, *F3'H2*) were highly homologous (sequence divergence: 23.44% across all samples; Figure S2) whereas *F3'H3* was markedly divergent from the former two (Figure S2). Similarly, *DFR3* is highly divergent from the other two copies (i.e., *DFR1* and *DFR2*). This marked divergence in *F3'H3* and *DFR3* coupled with more widespread expression of these two copies in leaf tissues across species (compared to expression of *F3'H1*, *DFR1*, *DFR2*, and to a lesser extent *F3'H2* in petal tissues; Fig. 2) suggests potential functional diversity of these two enzymes.

Using qRT-PCR, developmental stage-specific expression of multiple copies of *F3'H* and *DFR* found in *R. simplex* petals (i.e., *DFR1*, *DFR2*, *DFR3* and *F3'H2* and *F3'H3* [*F3'H1* not expressed in petals of this species; Fig. 2]) confirmed transcriptome-based expression patterns of ABP enzymes in this species (Fig. 3). These data further support the conclusion that our transcriptome assemblies are relatively complete and are of high quality. Analyses of *R. simplex* indicated that expression of *F3'H2*, *F3'H3* and *DFR3* all peaked at stage two of five developmental stages, also suggesting potentially coordinated regulation of those three genes (Fig. 3).

Network based co-expression analysis across tissue types in all samples recovered strong, correlated expression among *CHS*, *F3H*, *DFR1*, and *ANS* as well as between *CHI* and *DFR3* ( $r \geq 0.65$ ), suggesting possible blocks of structural genes under regulatory control (Fig. 4). Taken together, results indicate that regulation of ABP is highly coordinated and likely involves different regulatory blocks.

## Identification and co-expression analysis of candidate ABP regulatory genes

Our homology-based searches for candidate MBW regulatory elements in *Ruellia* recovered four clades of MYBs (Fig. 5), two of WDRs (Figure S3), and three of bHLHs (Figure S4). Transcript expression indicated that the three families of transcription factors exhibited different expression patterns: expression of all four clades of MYBs was specific to petals (Fig. 5) whereas WDRs and bHLHs were expressed in both petal and leaf tissues (Figures S3 and S4). Of the four MYB clades, copies of *MYB10L1* were relatively highly expressed compared to copies in the *MYB10L2*, *MYB10L3*, and *MYB10L4* clades (Fig. 5B). *MYB10L3* and *MYB10L4* were expressed only in a few species and at relatively low levels (Fig. 5).

Network based co-expression analysis across tissue types in all samples recovered strong, correlated expression of *MYB10L1* with candidate structural genes: *CHS*, *F3H*, *DFR1*, *DFR2* and *ANS* ( $r \geq 0.65$ ; Fig. 4) and a weak correlation with *F3'5'H* ( $r = 0.56$ ). No correlations among expression of *MYB10L2*, WDRs, bHLHs, and candidate structural genes were found; co-expression analyses were not conducted for *MYB10L3* and *MYB10L4* due to limited data.

We compared amino acid predicted sequences of MYBs from our four identified clades in *Ruellia* to amino acid translations of functionally validated MYBs known to regulate ABP in other flowering plants: *Arabidopsis thaliana*, *Petunia x hybrida*, *Pyrus pyrifolia*, and *Mimulus guttatus*. Specific attention was given to the amino acid sequence residue 'ANDV' because this residue has been reported to distinguish anthocyanin from non-anthocyanin MYBs (Lin-Wang et al., 2010). We found this residue to be present in candidate *Ruellia* MYBs from the MYB10L1 and MYB10L4 clades (Fig. 5B; note that expression levels of *MYB10L4* copies were markedly lower than were expression levels of *MYB10L1* copies). In contrast, replacement of the amino acid V with an I (resulting in the residue 'ANDI' and considered to be a conserved signature for non-anthocyanin gene regulation) characterized most of the *MYB10L2*s and *MYB10L3*s (as well as two of the four *Mimulus guttatus* sequences; Fig. 5B). We detected an additional non-synonymous mutation in this residue in *Ruellia lutea*, which resulted in an A to and S amino acid substitution (yielding 'SNDV'; Fig. 5B).

## Estimating amino acid substitutions in ABP loci and potential effects on protein function

To test whether codon sites of identified ABP loci are have potentially undergone adaptive evolution, we conducted tests of non-synonymous to synonymous substitution rates (dN/dS ratios, or  $\omega$ ) across all candidate structural genes plus one candidate MYB predicted to be significantly associated with the ABP in *Ruellia* (see Fig. 4). First, we implemented the site test model M0 in codeml, which assumes a single  $\omega$  ratio across all sites in an alignment. Results yielded dN/dS ratios that ranged between 0.20 and 0.72 (Table 2). Among all loci tested, *DFR3* had the highest dN/dS ratio (0.72) followed by *F3'H2* (0.54) and *MYB10L* (0.52). The fact that no gene was found to have dN/dS ratio > 1 indicates that all ABP genes are predominantly under purifying selection. However, because positive selection is unlikely to affect all sites in a given protein over extended evolutionary time, we further explored the potential of positive selection on individual amino acid sites. Results indicated that several sites in both *DFR3* and *F3'H2* are likely to be under positive selection (Fig. 6). Subsequent protein secondary structure analysis indicated that four of the sites that appear to be under positive selection in *F3'H2* and five sites in *DFR3* are located inside protein binding regions (PBRs).

Table 2  
Table showing dN/dS ratios of candidate ABP genes in *Ruellia*.

	<i>CHS</i>	<i>CHI</i>	<i>F3H</i>	<i>F3'H1</i>	<i>F3'H2</i>	<i>F3'H3</i>	<i>F3'5'H</i>	<i>DFR1</i>	<i>DFR2</i>	<i>DFR3</i>	<i>ANS</i>	<i>UF3GT</i>	<i>MYB10L</i>
dN	0.07	0.07	0.11	0.08	0.17	0.1	0.06	0.06	0.04	0.14	0.11	0.08	0.17
dS	0.25	0.28	0.25	0.17	0.32	0.27	0.19	0.18	0.2	0.2	0.38	0.22	0.33
dN/dS	0.29	0.25	0.45	0.5	0.54	0.38	0.31	0.36	0.2	0.72	0.29	0.36	0.52

We further examined the potential effects of amino acid substitutions on protein function at the above sites likely to be under positive selection. Using algorithms that predict both secondary structure and functional effects, we found one substitution in *F3'H2* and three substitutions in *DFR3* for which a moderate functional effect of the mutation is predicted (colored as pink in Fig. 6). These results indicate that, at these sites, observed amino acid substitutions are more likely to influence the stability or substrate binding affinity of the resulting protein or even cause protein dysfunction. At all other sites, observed amino acid substitutions colored as green indicate a strong signal for neutral effect.

## Co-expression analysis of all transcripts and associations with ABP loci

After removing transcripts that either had low expression levels or were expressed only in a small fraction of our study species (i.e., fewer than 3), a total of 60,613 transcripts were tested for correlated expression with the 13 ABP structural genes identified in *Ruellia*. Significant positive correlations were found for 1748 transcripts (2.89% of total transcripts present in 3 or more

species, Pearson's  $r \geq 0.65$ ,  $P < 0.05$ ). We used the KEGG classification system to predict and annotate functionality of these ABP-associated transcripts by assigning them to known biochemical pathways. Of the 1748 transcripts, 532 (30.43%) were assigned KEGG annotations. Among these, genes involved in a variety of metabolic pathways formed the largest group (Figures S5 and S6) among which 10 are putatively involved in phenylpropanoid biosynthesis and two are involved in flavone and flavone biosynthesis.

To gain further insight into interactions between transcripts and ABP loci as well as identify potential novel regulatory elements, transcription factors were further characterized using a custom plant transcription factor database derived from the Plant Transcription Factor Database. We identified 13 families of TFs that are positively correlated to ABP loci expression in *Ruellia* (Figure S5). Among these, MYBs were the dominant group of ABP-associated TFs and among 19 specific MYBs identified, nine belonged to the R2R3 family including *MYB10L1* that we earlier identified as likely to be especially relevant to ABP expression in *Ruellia* based on homology to functionally verified MYBs in other plant families (Fig. 5, Table S3). To further characterize potential function roles of the remaining eight R2R3 MYBs, we used amino acid sequences (derived from *R. simplex*) in blast searches against the NCBI nr database as well as in phylogenetic analysis with R2R3 MYBs from *A. thaliana* (Figure S7). Function annotations of these R2R3 MYBs suggest regulatory roles in either anthocyanin biosynthesis (*RsMYB10L1*, *RsMYB10L2*, *RsMYB10L3*, *RsMYB10L4* and *RsMYB114L*, homologs of *AtMYB75*, and *AtMYB114*; *RsMYB305L1* and *RsMYB305L2*, homologs of *AtMYB57*, *NtMYB305*; Borevitz et al., 2000; Liu and Thornburg, 2012) or lignin biosynthesis (*RsMYB46*, homolog of *AtMYB46*; *RsMYB306L*, homologs of *AtMYB31*; *RsMYB308L*, *RsMYB330L1* and *RsMYB330L2*, homologs of *AtMYB4*; Patzlaff et al., 2003; Newman, et al., 2004; Heppel et al. 2013; Cai et al., 2014; Wang, et al 2014; Galeano, et al., 2015; Agarwal et al., 2016; Koshiba et al., 2017).

## Discussion

In present work, we have studied the effect of mutations in ABP genes—coding and regulatory—and the effects of expression of these genes on flower color variation. Our results demonstrate the importance of both structural and regulatory changes during ABP evolution in *Ruellia*. We identified 12 candidate ABP structural genes plus 9 candidate transcriptional regulators, including 4 MYBs, 2 WDRs, and 3 bHLHs among the 10 investigated species. Through co-expression network analyses, we found evidence for strong correlated expression among most ABP structural genes as well as between most of these genes and one candidate regulatory MYB.

Phylogenetic relationships among the 10 species here investigated yield a hypothesis of four potential flower color transitions within the group: purple to red, red to yellow, red to purple, and purple to yellow (Fig. 1A). These four flower color transitions are consistent with patterns documented in a study with much more extensive phylogenetic sampling (Tripp and Manos 2008) and implicate both gain and loss of function mutations during flower color evolution in *Ruellia*. Prior studies on ABP have yielded evidence that four classes of mutations contribute importantly to color shifts across angiosperms: (1) mutations in ABP structural genes that cause either loss of function or changes to enzyme binding affinity, (2) mutations in cis-regulatory regions of ABP structural genes, and (3) mutations in coding regions or (4) cis-regions of ABP regulatory genes that impact expression of structural genes (Quattrocchio et al., 1999; Schwab et al., 2006; Whittal et al., 2006; Yakushiji et al., 2006; Chiu et al., 2010; Lin-Wang et al., 2010; Heppel et al., 2013; Shin et al., 2013; Medina-Puche et al., 2014; Yuan et al., 2014; Zhang et al., 2014). The purple to red color transition that occurred between the *R. breedlovei* lineage and the clade containing *R. fulgida* likely involved a loss of function mutation in the *F3'5'H* gene in the latter: our data showed that all essential ABP structural genes including *F3'5'H* itself were expressed at detectable levels (Fig. 2), but sequence analysis of *F3'5'H* revealed a premature stop codon caused by a C to T point mutation at position 373 in *R. fulgida* (Figure S8). By contrast, both our *de novo* and reference-based approaches failed to assemble *F3'5'H* in the red flower species *R. brevifolia*, which belongs to the same clade as *R. fulgida*, and expression analysis indicated *F3'5'H* was expressed at extremely low levels (Fig. 2). These data suggest a regulatory mutation was likely involved in this purple to red color transition. In the yellow-flowered species *R. bourgaei*, *R. speciosa*, and *R. lutea*, we observed down-regulation of multiple structural genes, including down-regulation of *F3H* in the closely related *R. bourgaei* and *R. speciosa* as well as down-regulation of *ANS* in all three species (Fig. 2). These results suggest either independent loss of functions in the cis-regulatory elements of these structural genes in the two clades or, more likely, mutations in shared

regulatory genes given that expression of *F3H* and *ANS* is strongly coordinately regulated (Fig. 4). Finally, the shift from red to purple flowers that occurred between *Ruellia elegans* and *R. hirsuto-glandulosa* likely involved reactivation of the *F3'5'H* pathway branch. Although restoration of pathway function is expected to be more difficult than degradation of function (Rausher, 2008, Sobel and Streisfeld, 2013), putative gains in floral anthocyanins have been documented (Armbruster, 2002; Kay *et al.*, 2005). In *R. elegans*, malvidin—a derivative of delphinidin—was detected in leaf tissues albeit in low concentrations, indicating functionality of the *F3'5'H* pathway in this species (Data S2). Restoration of the *F3'5'H* pathway in petals of *R. hirsuto-glandulosa* may have been achieved by mutation in the *cis*-regulatory region of this enzyme, leading to decreased binding affinity of transcription inhibitors or increased binding affinity of transcription activators.

With exception of a few mutations in *F3'H2* and *DFR3* (Fig. 6) and two mutations yielding premature stop codons in *F3'5'H* (Figure S8), sequence comparisons and protein predictions of ABP structural genes among species with different flower colors indicated that the majority of amino acid mutations likely had neutral or minimally different functional effects, thus unlikely to result in enzyme dysfunction (Data S3). Thus, much of the observed variation in flower colors and anthocyanin accumulation in *Ruellia* is more likely to be the result of mutations impacting the expression of ABP structural genes (Fig. 2). Regulatory mutations are considered to be among the most important factors in driving morphological evolution (Barrier *et al.*, 2001; Whittall *et al.*, 2006) given that such changes can potentially lead to modification in expression of entire pathways. Through co-expression analysis, we found that expression of one of the R2R3 MYBs identified in our study—*MYB10L1*, which is homologous to the known ABP regulator *MdMYB10* in apples (*Malus × domestica*; Wang *et al.*, 2010)—was highly associated with several structural genes including *CHS*, *F3H*, *DFR1*, *DFR2*, and *ANS* (Fig. 4). Thus, mutations that affect *MYB10L1*, be they regulatory or coding, may potentially impact all five of these candidate structural genes. However, sequence comparison of *MYB10L1* across all species of *Ruellia* sampled here recovered no lethal mutations (Fig. 5, Data S3), suggesting *MYB10L1* itself is likely regulated by a higher order of regulators.

Notably, based on patterns of expression correlation with *MYB10L1*, the regulation of candidate ABP structural genes in our dataset can be separated into two distinct regulatory blocks (Fig. 4). Group 1 contains genes whose expression are likely under the control of *MYB10L1*: *CHS*, *F3H*, *F3'5'H* (significantly correlated with *DFR2*,  $r = 0.65$ , weakly associated with *MYB10L1*,  $r = 0.56$ ), *DFR1*, *DFR2*, and *ANS*. Group 2 contains genes whose expression is independent of *MYB10L1*, among which *CHI* and *DFR3* seem to be regulated in a coordinated manner (Fig. 4). In contrast to these two blocks, each copy of *F3'H* and the single copy *UF3GT* are regulated independently. While independent regulation of *F3'H* has been observed in other studies (Jeremy *et al.*, 2015), co-expression of *UF3GT* with *CHS* or *F3H* in *Arabidopsis* has been reported in the past (Shin *et al.*, 2013; Ali and McNear, 2014). Thus, the observed expression pattern of *UF3GT* in the present study seems to be a unique feature of our dataset.

Prior research has demonstrated lower rates of non-synonymous mutations among genes upstream in the ABP pathway compared to downstream genes (Rausher *et al.*, 1999), likely as a function of relaxed evolutionary constraint on downstream (vs. upstream) genes (Lu and Rausher, 2003) or high rates of positive selection on downstream genes, potentially playing a role in adaptive floral color evolution (Huang *et al.*, 2016). This pattern is expected given that upstream enzymes function in a much broader set of biosynthetic pathways (e.g., lignin and other flavonoids) than downstream enzymes, which are largely specific to anthocyanin production (Figure S9). Data from *Ruellia* suggest similar patterns of rate evolution of ABP genes compared to prior studies. Specifically, two of the lowest dn/ds substitution rates were documented for *CHS* and *CHI* (Table 2), consistent with a hypothesis that these two genes are undergoing stronger purifying selection. Furthermore, across all candidate ABP loci, loss of expression was observed only in instances of genes downstream from *CHS* and *CHI* (Fig. 2), likely reflecting the essential role of these two enzymes in other metabolic pathways besides the ABP (Winkel-Shirley, 2001).

Flower color often reflects adaptation to pollinators (Bradshaw and Schemske, 2003). For example, floral anthocyanin content has a direct effect on pollinator behavior (Schemske and Bradshaw, 1999) and plant fitness (Schemske and Bierzychudek, 1999; Bradshaw and Schemske, 2003). That all candidate structural genes identified in present study are under purifying selection adds to evidence that selection acts to maintain overall primary protein function of ABP genes. However, our data also yielded evidence that selection may drive diversification of ABP products in some lineages; specifically, we found signatures of positive selection in *DFR* and *F3H* across species (Fig. 6), and some of these sites are located in or nearby protein binding

regions (PBR). This is especially the case for *DFR3*, in which five out of seven sites under positive selection are located within predicted PBS. Mutations occurring at these sites are more likely to impact substrate binding affinities of the corresponding enzyme resulting in new specific interactions (Reva et al., 2011). In this study, we found multiple, orthologous copies of both *F3H* and *DFR* to be present and here hypothesizes that duplication of these genes may have facilitated functional diversification of the ABP in *Ruellia* (Huang et al., 2016).

Transcription factors (TFs) are key players in regulating flux through secondary metabolic pathways by controlling relative levels of gene expression (Broun, 2005). To further investigate regulation of the ABP in the context of potential interactions with other metabolic pathways, we conducted a phylogenetic study of candidate MYBs as well as co-expression analyses of broader diversity of TFs across all species. We identified nine candidate R2R3 MYBs associated with the ABP identified in *Ruellia* (Table S3). Among these, orthologs of *RsMYB10L1*, *RsMYB114L* (i.e., *AtMYB75*; Borevitz et al., 2000), *RsMYB305L1*, and *RsMYB305L2* (i.e., *NtMYB305*; Wang et al., 2014) have been reported to function in regulating anthocyanin accumulation. Our investigation demonstrated that *RsMYB10L1* expression is highly correlated with five ABP structural genes (Fig. 4), and phylogenetic analyses demonstrated that *RsMYB10L1* belongs to a strongly supported clade that contains *AtMYB75* orthologs (Figure S7). This result in combination with sequence analysis and other expression patterns (Fig. 5) suggest specifically that *RsMYB10L1* is a very strong candidate regulator of the ABP in *Ruellia*. Pilot experimentation with virus-induced gene silencing focusing on ABP knockdowns of *RsMYB10L1* have yielded additional support for this candidate (Y. Zhuang and E. Tripp, unpub. data; Figure S10). Our phylogenetic analyses placed the remaining *Ruellia* R2R3 MYBs in clade with *Arabidopsis* orthologs that function in lignin biosynthesis (e.g., *AtMYB4*, *AtMYB31*, and *AtMYB46/AtMYB83*; Patzlaff et al., 2003; Newman, et al., 2004; Cai et al., 2014; Agarwal et al., 2016; Koshiba et al., 2017). These results support a hypothesis of different functional roles of the R2R3 MYBs recovered in *Ruellia*—some in ABP production vs. others potentially in lignin production—albeit in pathways that interact via common upstream enzymes (Figure S9; Tohge et al., 2005). Functionally, anthocyanins and lignins are both proposed to protect plants against UV radiation (Mouradov et al., 2014), and redirection of metabolic fluxes between the lignin and flavonoid pathways has been observed in many plant species (Ring et al., 2013). Over-expression of *AmMYB308* in transgenic tobacco plants and several of its orthologs belonging to the same clade led to significant repression of lignin biosynthesis and increased total flavonoids (Tamagnone et al., 1998; Ma et al., 2011). Other studies have similarly demonstrated MYBs as negative regulators of lignin content (Agarwal et al., 2016) whereas others likely serve as activators of lignin biosynthesis (Bevan et al., 2003; Zhao and Bartley, 2014; Koshiba et al., 2017). Taken together, data suggest extensive interactions between anthocyanin and lignin biosynthesis through modification of expression of relevant TFs. As such, future attempts to understand regulatory impacts on flower color evolution in *Ruellia* would optimally incorporate information from other biosynthetic pathways that likely interact with anthocyanin production.

We also identified several TFs that are highly co-expressed with ABP-associated TFs in *Ruellia*. These include auxin responsive factors (ARFs), abscisic acid signaling genes (ABA), and other TFs, some of which function as activators and others as repressors based on prior studies in other plants (Figure S5; Ulmasov et al., 1999; Okushima et al., 2007; Gutierrez et al., 2009; Wheeler et al., 2009; Daminato et al., 2013;). In present study, we found positive co-expression between two auxin signaling pathway activators *ARF6* and *ARF19* (Ulmasov et al., 1999; Okushima et al., 2007) and ABP-associated TFs in *Ruellia*. We speculate that this positive association results from the role of auxin in flower development. Specifically, the accumulation of anthocyanins gradually increases during flower development in *Ruellia* (shown in Fig. 3A). Auxin is required for the initiation of floral primordia and disruption of auxin biosynthesis leads to the failure of flower formation (Cheng et al., 2007). Our hypothesis is further supported by the observed correlation of a group of MADS-box TFs with ABP structural genes, which is exclusively expressed only in petal tissues (Figure S11). Sequence analysis indicated one gene was homologous to *AGL6*, two genes were homologous to *MADS1*, and another two genes were homologous to *MADS2*, all of which impact flower development (Ma, 1994; Schwab et al., 2006). Taken together, our data suggested strong, coordinated regulation between anthocyanin accumulation and flower development.

## Conclusions

The ABP is one of the most well-studied secondary metabolic pathways in plants. Although key structural genes have been extensively characterized and are highly conserved across unrelated species (Petroni and Tonelli, 2011), the complex mechanisms regulating the ABP are still not fully understood largely as a result of prior emphasis on model plant systems. The present study provides the first steps to understanding molecular evolution and expression of ABP genes in *Ruellia*. Specifically, it yields a first comprehensive survey of candidate structural and regulatory loci involved in the production of diverse flower colors in the genus. Data from our expression and co-expression analyses indicate that regulatory mutations have likely been more important in color shifts but that mutations to structural genes have also played a role. Finally, different portions of the ABP appear to be under coordinated regulation such that a single 'regulatory block' is unlikely to characterize the entire pathway, perhaps as a function of extensive interactions with other metabolic pathways.

## Methods

### Plant materials

This study sampled flower petal (P) and/or leaf (L) tissues from the following species, all of which are in cultivation in glasshouses at the University of Colorado–Boulder: *Ruellia bourgaei* (P), *R. breedlovei* (P), *R. brevifolia* (P), *R. elegans* (P, L), *R. fulgida* (P, L), *R. hirsuto-glandulosa* (P, L), *R. longepetiolata* (P), *R. lutea* (P, L), *R. simplex* (P, L), and *R. speciosa* (P, L). Flower petal tissues were removed from S3 stage (Fig. 3) and leaf tissues were collected at mature stage, ca. 2–3 nodes below the uppermost meristem. All species except *R. longepetiolata* and *R. elegans* were derived from wild field populations; the former two species were acquired and grown from cuttings sent by colleagues.

### Anthocyanidin identification

We followed anthocyanidin identification and quantification methods described in Zhuang and Tripp (2017). Briefly, 25 mg of silica-dried leaf or petal tissue was placed in 1.5 ml tubes filled with 2N HCl. Sugars were cleaved from anthocyanin molecules by placing tubes in a 103 °C heat block for 90 minutes then centrifuged for 5 min at 10,000 rpm at room temperature to obtain supernatants. Retained supernatants were washed twice with 400  $\mu$ l of ethyl acetate and centrifuged for 1 min at 10,000 rpm to restore phase separation. The pigmented bottom layer was further washed twice with 200  $\mu$ l of isoamyl alcohol and injected into an Agilent 1260 Infinity system (Thermo Scientific) for anthocyanidin identification and quantification.

### cDNA library construction and sequencing

Fresh leaf and petal tissues were collected from plants growing in the glasshouses and placed immediately into liquid N<sub>2</sub> for RNA extraction. Total RNA was extracted with MasterPure™ RNA Purification Kits (Epicentre) following the manufacturer's protocols. Integrity of total RNA was determined by running samples on a 1.2% agarose gel. RNA-seq library preparation was conducted using an Illumina ScriptSeq Complete Kit, Low Input (BL1224, Illumina) according to manufacturer's instructions. The sequence-ready libraries were quantified using a Qubit (Invitrogen) and quality was assessed using a Bioanalyzer. Libraries were sent to the Genomics and Microarray Core, University of Colorado–Anschutz Medical Campus then sequenced on an Illumina HiSeq2500 using 2 × 125 bp paired-end (PE) chemistry. Downstream analysis of the resulting raw data is described in Figure S1. Resulting sequences have been deposited at NCBI: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA323650>.

### Raw reads processing

We used Trimmomatic (Bolger et al., 2014; LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) to remove low quality bases and adaptor contaminations in our raw reads. After quality filtering, reads were then error-corrected using SEECER, a hidden Markov model (HMM)-based probabilistic error correction tool, with default parameters (Le et al., 2013).

### Hybrid RAD loci–transcriptome approach to reconstructing a species phylogeny

To understand evolutionary relationships among the ten species used in our study, we developed a simple pipeline that uses unassembled transcriptome reads for phylogenetic reconstruction. First, all error-corrected reads were screened for the presence of the sequence tag "GAATTC", which represents the *EcoRI* enzyme recognition sequence. The sequence tag 5'-AATTC as well

as its downstream sequences were extracted. The resulting reads were used for *de novo* locus assembly implemented in pyRAD v3.0.6 (Eaton, 2014; `-mindepth 15, -clustering threshold 0.85, -minsamples: 3`). Resulting biallelic loci were extracted from the final alignment using an in-house Perl script. We used maximum likelihood methods to reconstruct phylogenetic relationships in RAxML v8 (Stamatakis, 2014; `-f a; -m GTRGAMMA; -N 200`).

## Transcriptome de novo assembly and quality assessment

Tissue-specific transcriptomes were assembled *de novo* using Trinity software release v2.2.0 (Grabherr et al., 2011) with a minimum contig length cutoff of 200 bp. Different parameter combinations were tested to obtain the best assembly (`-min_kmer_cov: 1 or 2; -normalize_max_read_cov 30, 40 or 50; -jaccard_clip on or off`). Assembly quality was evaluated using TransRate v1.0.125 (Smith-Unna et al., 2016). Assembled transcriptomes with the best TransRate scores were selected for ORF identification using TransDecoder (<https://github.com/TransDecoder/TransDecoder>). Identified ORF sequences were then clustered using CD-HIT (Li and Godzik, 2006) to remove redundant sequences at a 99% similarity threshold. Mapping metrics were obtained using bowtie2 v2.2.6 (Langmead and Salzberg, 2012) by mapping error-corrected reads back to their corresponding tissue-specific assembly. The completeness of the TransRate-selected transcriptome was assessed using BUSCO v2.0.1 (Simão et al., 2015; plant BUSCO dataset downloaded 3/31/17).

## PCA analysis

To generate a uniform read count table for downstream analysis, identified ORF sequence files from all samples were pooled together and clustered using CD-HIT-EST with a similarity threshold of 95%, and each resulting transcript was considered as a unigene then used as a reference for reads mapping. Reads from each tissue-specific library were mapped to the constructed reference using bowtie2 (`-sensitive`), and eXpress v1.5.1 (Roberts and Pachter, 2013) was used to extract raw read counts and FPKM value for each sample. Transcripts were included in downstream analysis when > 10 reads were recovered in a minimum of five samples. We then conducted a regularized logarithm (rlog) transformation using the rlog function in DESeq2 (Love et al., 2014) on the remaining data. A PCA analysis was conducted on these transformed data using the plotPCA function in DESeq2 and results were visualized using the R package ggplot2 (Wickham, 2011).

## Identification of ABP structural genes

To identify candidate structural genes involved in the ABP in *Ruellia*, six core pathway genes from *Arabidopsis* (*CHS*: AT5G13930, *CHI*: AT3G55120, *F3H*: AT3G51240, *DFR*: AT5G42800, *ANS*: AT4G22880, and *UF3GT*: AT5G54060) and two branching genes from *Penstemon neomexicanus* (*F3'H*: KM388824 and *F3'5'H*: KM388828) were used as references for homolog identification using NCBI-BLAST (Altschul et al., 1997), with a blast e-value threshold of  $1 \times 10^{-100}$  for both blastn and blastp. To further improve the quality of this dataset, all identified transcripts were further blasted against NCBI's nr Database; candidate transcripts were retained only if the top blast hit was annotated as a structure gene belonging to the ABP. cDNA sequences of transcripts obtained from all species were aligned to their corresponding reference genes to assess sequence integrity using MAFFT v7.3.10 (Katoh *et al.*, 2012; `mafft -maxiterate 1000 -localpair`).

For species for which we were unable to assemble full-length cDNAs for all eight ABP loci, a second round of *de novo* assembly was conducted. First, all error-corrected reads were mapped to the reference gene of its most closely related species (see Fig. 1A) using bowtie2 v2.2.6 (Langmead and Salzberg, 2012; `-sensitive`). All mapped reads were then extracted using samtools v1.4.1 (Li et al., 2009) and served as input for *de novo* assembly in SPAdes v3.9.0 (Bankevich et al., 2012) with default parameters.

## Identification of conserved MYBs, bHLH and WDR

To identify candidate ABP regulatory MYB, bHLH, and WDR-type transcription factors, all protein sequences translated by TransDecoder were subject to domain searching for the MYB domain (PS51294), bHLH domain (PS50888) and WD-40 repeat (PS50294) using ps\_scan (Gattiker et al., 2002) with default parameters. For MYB domain-containing proteins, the number of MYB domains was counted using an in-house Perl script; we retained only R2R3 type MYBs for further analysis. Seven MYBs with a known or candidate role in regulating the ABP were used as references during searches for homologs in *Ruellia* using NCBI blastp (Altschul et al., 1997) with a blast e-value threshold of  $1 \times 10^{-100}$ . These included 8 R2R3 type MYB proteins:

PAP1 (AtMYB75, AT1G56650), PAP2 (AtMYB90, AT1G66390), mgv1a023671m, mgv1a024996m, mgv1a019326m, mgv1a025765m; 10 bHLH proteins: AtbHLH34 (AT3G23210), GL3 (AT5G41315), EGL3 (AT1G63650), TT8 (AT4G09820), AtbHLH32 (AT3G25710) MYC3 (AT5G46760), MYC4 (AT4G17880), Migut.A00733.1 Migut.D00287.1 Migut.E01090.1; and two WDR proteins: TTG1 (AT5G24520) and ATAN11 (AT1G12910).

## Co-expression analysis of candidate ABP genes

To evaluate the expression correlations among identified candidate ABP genes, their FPKM values were used for testing the significance of correlations (an FDR adjusted p-value threshold of  $< 0.05$ ) was testing using the `corr.test()` function in the R package "psych" (Revelle and Zinbarg, 2009) and p-values were adjusted using the Holm-Bonferroni method. All statistical analyses were conducted within the R environment. The correlation was visualized as a matrix using R package "corrplot" (Wei and Simko, 2013). For candidate ABP structural genes associated with *MYB10L1*, correlations were predicted using a linear model with a 95% confidence level interval (`geom_smooth(method="lm", level = 0.95)`) and visualized using ggplot2 (Wickham, 2011).

## Non-synonymous to synonymous mutation rates and tests for selection

To test for evidence of positive selection on ABP genes, MAFFT -aligned cDNA sequences and their corresponding coding sequences were used to guide coding sequence (CDS) alignments in Pal2Nal (Suyama et al., 2006; <http://www.bork.embl.de/pal2nal/>). The ratio of non-synonymous to synonymous mutations (dN/dS, or  $\omega$ ) was estimated using codon-based maximum likelihood methods implemented in codeml, a module in the PAML package (version 4.8; Yang, 2007). To test for positive selection across an entire gene, we used the M0 model, which assumes the same  $\omega$  ratio for all sites in the gene. To test for positive selection in a site-specific manner, we first identified sites under positive selection through Naive Empirical Bayes (NEB) analysis ( $P > 95\%$ ) using the site models M1, M2, M7 and M8, which allow the  $\omega$  ratio to vary among sites. Models M1 + M2, and M7 + M8, form two pairs of models that can be used to test for positive selection using a likelihood ratio test (LRT). Model M1 and M2 allow the site classes  $\omega = 1$  and  $0 < \omega < 1$ , and model M2 additionally includes a third class:  $\omega > 1$  (Yang, 2007). Models M7 and M8 include several site classes with  $\omega$  ratios that follow a beta distribution, and model M8 additionally includes a  $\omega > 1$  class. Two sets of model comparisons were conducted for each gene: the first compares M1/M2 whereas the second compares M7/M8. Codon sites that are under positive selection were identified using the Bayes Empirical Bayes (BEB) method (Yang, 2007).

## Secondary structure analysis and functional predictions

To examine potential effects of amino substitutions on protein function, we predicted the secondary structure of proteins as well as functional effects of mutations at amino acid sites under positive selection. Secondary structure of proteins (helices, sheets, and coils) was predicted using the PredictProtein online server (Yachdav et al., 2014; <https://www.predictprotein.org>, last accessed 07/21/2017). Functional effects of sequence variants were predicted using snap2, which is integrated into the predictProtein server (Yachdav et al., 2014).

## Identification of ABP interactive pathways through co-expression network analysis

To test for correlations in expression among assembled transcripts and ABP genes, transcripts with at least 10 reads counts and transcripts that were found in at least three samples were selected. Pearson correlation coefficients were calculated using the `corr.test` function in the R package "psych" (Revelle and Zinbarg, 2009). To calculate adjusted P values, expression levels of each ABP gene were permuted 1000 times and calculated as  $(r+1)/(n+1)$ , where  $n = 1000$  and  $r$  is the number of these replicates that produce a test statistic greater than or equal to that calculated for the actual data (North et al., 2002). We deemed correlations to be significant when Pearson's  $r \geq 0.65$  and  $P < 0.05$ . All transcripts with expression levels significantly associated with the 12 identified candidate structural genes were annotated using Trinotate pipeline release version 3.0.0 (Grabherr et al., 2011; <http://trinotate.sourceforge.net/>) with a blast e-value threshold of  $1 \times 10^{-5}$ . We additionally built a custom plant transcription factor database for ABP loci using data downloaded from the Plant Transcription Factor Database v4.0 (Pérez-Rodríguez et al., 2010; <http://planttfdb.cbi.pku.edu.cn>, last accessed 04/18/2017). KEGG pathway analysis was

conducted on all resulting transcripts using BlastKOALA (Kanehisa et al., 2016; <http://www.kegg.jp/blastkoala/> Last accessed 4/01/2017) and was further processed using an in-house Perl script then visualized using the R package ggplot2 (Wickham, 2011).

## Phylogenetic analysis and sequence comparison

We aligned translated protein sequences of ABP genes using MAFFT v7.3.10 (Katoh *et al.*, 2012; `mafft -maxiterate 1000 -localpair`) and then identified conserved regions of resulting alignments using Gblocks (Talavera and Castresana, 2007). Phylogenetic analyses were conducted on final alignments using RAxML and the GTR + GAMMA substitution model with 200 bootstrap replicates. Resulting phylogenetic trees were further edited and viewed using Evolview (Zhang et al., 2012; <http://www.evolgenius.info/evolview>, last accessed 3/21/2017).

## Real time qPCR

To confirm transcripts abundance measured from the RNA-Seq data and better characterize assembled homologs of *F3'H* and *DFR*, we extracted total RNA from petal tissues of *Ruellia simplex* collected at different growth stages using MasterPure™ RNA Purification Kit (Epicentre) following the manufacturer's protocols. For reverse transcription, ~ 500 ng of DNase I-treated total RNA was used for cDNA synthesis using the GoScript™ Reverse Transcriptase system (Promega) with the oligo(dT)<sub>15</sub> primer. qPCR was carried out using SYBR green I based master Mix (Roche Life Science, Indianapolis, IN) with 2 min at 95 °C, 40 cycles of 20 s at 95 °C, 45 s at 57 °C, and 30 s at 72 °C, and then 5 min at 72 °C (Applied Biosystems). For gene expression analysis, the 18 s reference gene was used as an internal control to normalize *Ct* values. For each sample, three technical replicates were conducted.

## Declarations

### ***Ethics approval and consent to participate.***

Not applicable.

### ***Consent for publication.***

Not applicable.

### ***Availability of data and materials.***

The datasets generated and analyzed during the present study are available at Dryad (Study Number XXXXXX).

### ***Competing interests.***

The authors declare no competing interests.

### ***Funding.***

This work was supported by National Science Foundation's Division of Environmental Biology (Awards #1354963 and #1355138) to Erin Tripp and Lucinda McDade.

### ***Authors' contributions.***

YZ and ET generated, analyzed, and interpreted the data. YZ and ET co-wrote the manuscript. YZ and ET conceptualized the study.

### ***Acknowledgements.***

We thank Matt Schreiber and Heather Stone for conducting HPLC analyses.

## References

1. Ali, M.B. and McNear, D.H. Induced transcriptional profiling of phenylpropanoid pathway genes increased flavonoid and lignin content in *Arabidopsis* leaves in response to microbial products. *BMC Plant Biology*. 2014;14:84.
2. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997;25:3389-3402.
3. Agarwal, T., Grotewold, E., Doseff, A. I., and Gray, J. MYB31/MYB42 syntelogs exhibit divergent regulation of phenylpropanoid genes in maize, sorghum and rice. *Scientific Reports*. 2016;6:28502.
4. Armbruster W.S. Can indirect selection and genetic context contribute to trait diversification? A transition probability study of blossom-colour in two genera. *Journal of Evolutionary Biology*. 2002;15:468–486.
5. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S. and Pribelski, A.D. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 2012;19:455-477.
6. Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B. and Lepiniec, L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*. 2004;39:366-380.
7. Barrier, M., Robichaux, R.H. and Purugganan, M.D. Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences*. 2001; 98:10208-10213.
8. Bolger, A.M., Lohse, M. and Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
9. Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*. 2000;12:2383-2393.
10. Bradshaw, H. and Schemske, D.W. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*. 2003;426:176-178.
11. Bevan, M., Manseld, S., Whetten, R., Sederoff, R. and Campbell, M. Characterisation of a pine MYB that regulates lignification. *The Plant Journal*. 2003;36:743-754.
12. Broun, P. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. *Current Opinion in Plant Biology*. 2005;8:272-279.
13. Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *THE Plant Cell*. 2012; 24:1242-1255
14. Cai, B., Li, C.H. and Huang, J. Systematic identification of cell-wall related genes in *Populus* based on analysis of functional modules in co-expression network. *PloS one*. 2014;9: e95176.
15. Cheng, Y. and Zhao, Y. A role for auxin in flower development. *Journal of Integrative Plant Biology*. 2007;49:99-104.
16. Chiu, L.W., Zhou, X., Burke, S., Wu, X., Prior, R.L. and Li, L. The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiology*. 2010;154:1470-1480.
17. Daminato, M., Guzzo, F. and Casadoro, G. A SHATTERPROOF-like gene controls ripening in non-climacteric strawberries, and auxin and abscisic acid antagonistically affect its expression. *Journal of Experimental Botany*. 2013;64:3775-3786.
18. Eaton, D.A. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*. 2014;62:689–706.
19. Espley, R.V., Hellens, R.P., Putterill, J., Stevenson, D.E., Kutty-Amma, S. and Allan, A.C. Red coloration in apple fruit is due to the activity of the MYB transcription factor, *MdMYB10*. *The Plant Journal*. 2007;49:414-427.
20. Galeano, E., Vasconcelos, T.S., Vidal, M., Mejia-Guerra, M.K. and Carrer, H. Large-scale transcriptional profiling of lignified tissues in *Tectona grandis*. *BMC Plant Biology*. 2015;15:221.
21. Gattiker, A., Gasteiger, E. and Bairoch, A.M. ScanProsite: a reference implementation of a PROSITE scanning tool. *Applied Bioinformatics*. 2002; 1:107-108.

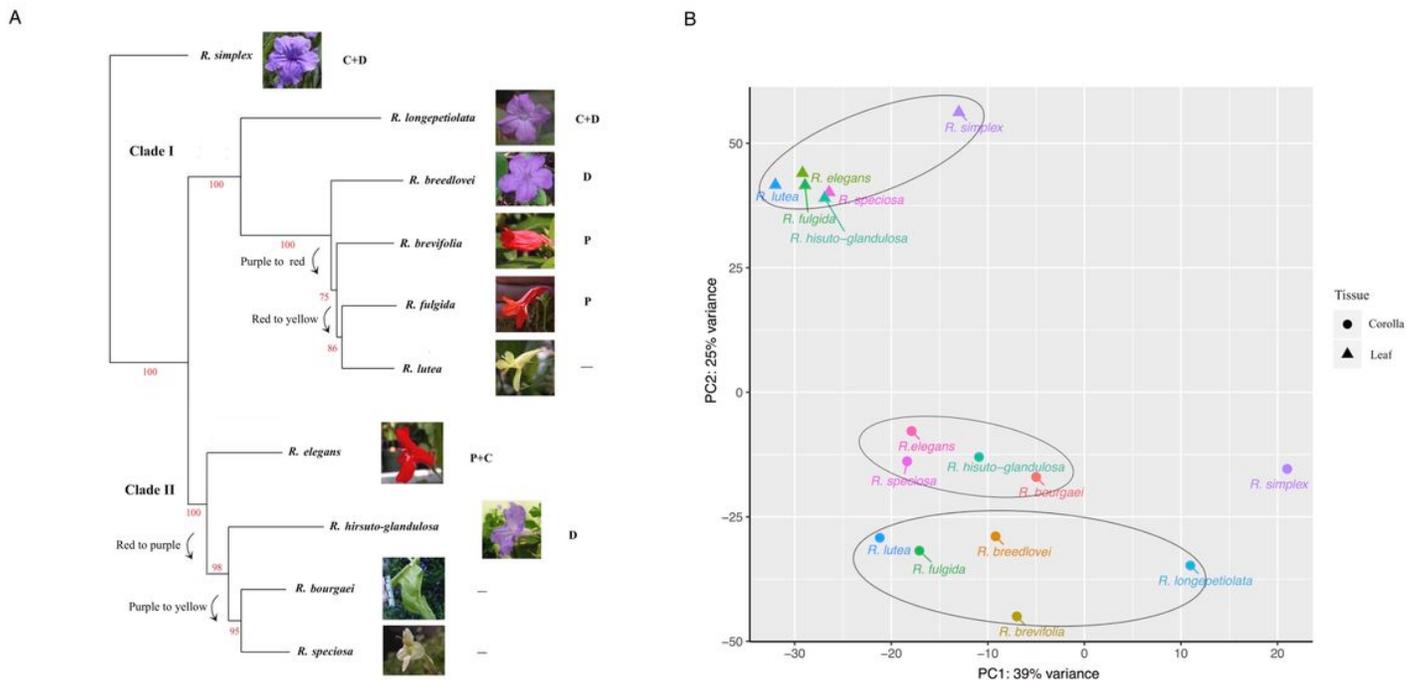
22. Goodrich, J., Carpenter, R. and Coen, E.S. A common gene regulates pigmentation pattern in diverse plant species. *Cell*. 1992;68:955-964.
23. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*. 2011;29: 644.
24. Gutierrez, L., Bussell, J.D., Păcurar, D.I., Schwambach, J., Păcurar, M. and Bellini, C. Phenotypic plasticity of adventitious rooting in *Arabidopsis* is controlled by complex regulation of AUXIN RESPONSE FACTOR transcripts and microRNA abundance. *The Plant Cell*. 2009;21:3119-3132.
25. Heppel, S.C., Jaffé, F.W., Takos, A.M., Schellmann, S., Rausch, T., Walker, A.R. and Bogs, J. Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Molecular Biology*. 2013; 82:457-471.
26. Hittinger, C.T., Johnston, M., Tossberg, J.T. and Rokas, A. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences*. 2010; 107:1476-1481.
27. Huang, B.H., Chen, Y.W., Huang, C.L., Gao, J. and Liao, P.C. Imbalanced positive selection maintains the functional divergence of duplicated *DIHYDROKAEMPFEROL 4-REDUCTASE* genes. *Scientific Reports*. 2016; 6:39031.
28. Kanehisa, M., Sato, Y. and Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*. 2016;428:726-731.
29. Katoh, K., Misawa, K., Kuma, K.i. and Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 2002; 30:3059-3066.
30. Koshiha, T., Yamamoto, N., Tobimatsu, Y., Yamamura, M., Suzuki, S., Hattori, T., Mukai, M., Noda, S., Shibata, D. and Sakamoto, M. MYB-mediated up-regulation of lignin biosynthesis in *Oryza sativa* towards biomass refinery. *Plant Biotechnology*. 2017;34:7-15.
31. Langmead, B. and Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;9:357-359.
32. Le, H.-S., Schulz, M.H., McCauley, B.M., Hinman, V.F. and Bar-Joseph, Z. Probabilistic error correction for RNA sequencing. *Nucleic acids research*. 2013;41:e109.
33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.
34. Li, W. and Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658-1659.
35. Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T.K., Espley, R.V., Hellens, R.P. and Allan, A.C. An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biology*. 2010;10:50.
36. Liu G, and Thornburg RW. Knockdown of *MYB305* disrupts nectary starch metabolism and floral nectar production. *The Plant Journal*. 2012;70:377–388
37. Love, M.I., Huber, W. and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014;15:550.
38. Lu, Y. and Rausher, M.D. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution*. 2003;20:1844-1853.
39. Ma, H. The unfolding drama of flower development: recent results from genetic and molecular analyses. *Genes and Development*. 1994;8:745-756.
40. Ma, Q.H., Wang, C. and Zhu, H.H. *TaMYB4* cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. *Biochimie*. 2011;93:1179-1186.
41. Medina-Puche, L., Cumplido-Laso, G., Amil-Ruiz, F., Hoffmann, T., Ring, L., Rodríguez-Franco, A., Caballero, J.L., Schwab, W., Muñoz-Blanco, J. and Blanco-Portales, R. *MYB10* plays a major role in the regulation of flavonoid/phenylpropanoid

- metabolism during ripening of *Fragaria* × *ananas* fruits. *Journal of Experimental Botany*. 2014;65:401-417.
42. Mouradov, A. and Spangenberg, G. Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Frontiers in Plant Science*. 2014;5:620.
43. Nesi, N., Jond, C., Debeaujon, I., Caboche, M. and Lepiniec, L. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *The Plant Cell*. 2001;13:2099-2114.
44. Newman, L.J., Perazza, D.E., Juda, L. and Campbell, M.M. Involvement of the R2R3-MYB, *AtMYB61*, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. *The Plant Journal*. 2004;37:239-250.
45. North, Bernard V., David Curtis, and Pak C. Sham. A note on the calculation of empirical P values from Monte Carlo procedures. *American Journal of Human Genetics*. 2002;71:439.
46. Okushima, Y., Fukaki, H., Onoda, M., Theologis, A. and Tasaka, M. *ARF7* and *ARF19* regulate lateral root formation via direct activation of *LBD/ASL* genes in *Arabidopsis*. *The Plant Cell*. 2007;19:118-130.
47. Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. *The Plant Cell*. 1999;11:1433-1444.
48. Patzlaff, A., McInnis, S., Courtenay, A., Surman, C., Newman, L.J., Smith, C., Bevan, M.W., Mansfield, S., Whetten, R.W. and Sederoff, R.R. Characterisation of a pine MYB that regulates lignification. *The Plant Journal*. 2003;36: 743-754.
49. Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corrêa, L.G.G., Rensing, S.A., Kersten, B. and Mueller-Roeber, B. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Research*. 2010;38:822-827
50. Petroni, K. and Tonelli, C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science*. 2011;181:219-229.
51. Rausher, M. D., R. E. Miller, and P. Tiffin. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* 1999;16:266-274.
52. Rausher, M.D. Evolutionary transitions in floral color. *International Journal of Plant Sciences*. 2008;169:7-21.
53. Reva, B., Antipin, Y. and Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*. 2011;39:118.
54. Revelle, W. and Zinbarg, R.E. Coefficients *alpha*, *beta*, *omega*, and the *glb*: Comments on Sijtsma. *Psychometrika*. 2009;74:145-154.
55. Ring, L., Yeh, S.-Y., Hücherig, S., Hoffmann, T., Blanco-Portales, R., Fouche, M., Villatoro, C., Denoyes, B., Monfort, A. and Caballero, J.L. Metabolic interaction between anthocyanin and lignin biosynthesis is associated with peroxidase FaPRX27 in strawberry fruit. *Plant Physiology*. 2013;163:43-60.
56. Roberts, A. and Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*. 2013;10:71-73.
57. Salisbury, B.A. and Kim, J. Ancestral state estimation and taxon sampling density. *Systematic Biology*. 2001;50:557-564.
58. Schemske, D.W. and Bradshaw, H.D. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences*. 1999;96:11910-11915.
59. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. and Weigel, D. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *The Plant Cell*. 2006;18:1121-1133.
60. Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K. and Martin, C. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *The Plant Cell*. 2006;18:831-851.
61. Shin, D.H., Choi, M., Kim, K., Bang, G., Cho, M., Choi, S.B., Choi, G. and Park, Y.I. *HY5* regulates anthocyanin biosynthesis by inducing the transcriptional activation of the *MYB75/PAP1* transcription factor in *Arabidopsis*. *FEBS letters*. 2013;587:1543-1547.
62. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–3212.

63. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M. and Kelly, S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*. 2016;26:1134-1144.
64. Sobel, J.M. and Streisfeld, M.A. Flower color as a model system for studies of plant evo-devo. *Frontiers in Plant Science*. 2013;10:236-242.
65. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312-1313.
66. Suyama, M., Torrents, D. and Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*. 2006;34:609-612.
67. Talavera, G. and Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*. 2007;56:564-577.
68. Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez-Macia, F.A., Roberts, K. and Martin, C. The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*. 1998;10:135-154.
69. Tanaka, Y., Sasaki, N. and Ohmiya, A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal*. 2008;54:733-749.
70. Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J.i., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B. and Kitayama, M. (Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *The Plant Journal*. 2005;42:218-235.
71. Tripp, E.A. and Manos, P.S. Is floral specialization an evolutionary dead-end? pollination system transitions in *Ruellia* (Acanthaceae). *Evolution*. 2008;62:1712-1737.
72. Tripp, E. A., and Tsai, Y. H. E. Disentangling geographical, biotic, and abiotic drivers of plant diversity in neotropical *Ruellia* (Acanthaceae). *PloS One*. 2017;12:e0176021.
73. Ulmasov, T., Hagen, G. and Guilfoyle, T.J. Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences*. 1999;96:5844-5849.
74. Wang, W., Liu, G., Niu, H., Timko, M.P. and Zhang, H. The F-box protein CO11 functions upstream of *MYB305* to regulate primary carbohydrate metabolism in tobacco (*Nicotiana tabacum* L. cv. TN90). *Journal of Experimental Botany*. 2014;65:2147-2160.
75. Wei, T. and Simko, V. corrplot: Visualization of a correlation matrix. R package version 0.73. 2013;230:11.
76. Weiss, M.R. Floral color change: a widespread functional convergence. *American Journal of Botany*. 1995;82:167-185.
77. Wen, J., Xiong, Z., Nie, Z.-L., Mao, L., Zhu, Y., Kan, X.-Z., Ickert-Bond, S.M., Gerrath, J., Zimmer, E.A. and Fang, X.-D. Transcriptome sequences resolve deep relationships of the grape family. *PLoS One*. 2013;8:e74394.
78. Wheeler, S., Loveys, B., Ford, C. and Davies, C. The relationship between the expression of abscisic acid biosynthesis genes, accumulation of abscisic acid and the promotion of *Vitis vinifera* L. berry ripening by abscisic acid. *Australian Journal of Grape and Wine Research*. 2009;15:195-204.
79. Whittall, J.B., Voelckel, C., Kliebenstein, D.J. and Hodges, S.A. Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology*. 2006;15:4645-4657.
80. Wickham, H. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2011;3:180-185.
81. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*. 2001;126:485-493.
82. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M. and Bernhofer, M. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Research*. 2014;42:337-343.
83. Yakushiji, H., Kobayashi, S., Goto-Yamamoto, N., Jeong, S.T., Sueta, T., Mitani, N. and Azuma, A. A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional *VvmybA1* allele. *Bioscience, Biotechnology, and Biochemistry*. 2006;70:1506-1508.

84. Yang, Ziheng. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007;24:1586-1591.
85. Yuan, Y.W., Sagawa, J.M., Frost, L., Vela, J.P. and Bradshaw, H.D. Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*). *New Phytologist*. 2014;204:1013-1027.
86. Zhang, H., Gao, S., Lercher, M.J., Hu, S. and Chen, W.H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Research*. 2012;40:569-572.
87. Zhang, Y., Cheng, Y., Ya, H., Xu, S. and Han, J. Transcriptome sequencing of purple corolla spot region in tree peony reveals differentially expressed anthocyanin structural genes. *Frontiers in Plant Science*. 2014;6:964-964.
88. Zhao, D. and Tao, J. Recent advances on the development and regulation of flower color in ornamental plants. *Frontiers in Plant Science*. 2015;6:261.
89. Zhao, L., Gao, L., Wang, H., Chen, X., Wang, Y., Yang, H., Wei, C., Wan, X. and Xia, T. The R2R3-MYB, bHLH, WDR, and related transcription factors in flavonoid biosynthesis. *Functional and Integrative Genomics*. 2013;13:75-98.
90. Zhao, X., Sheng, F., Zheng, J. and Liu, R. Composition and stability of anthocyanins from purple *Solanum tuberosum* and their protective influence on Cr (VI) targeted to bovine serum albumin. *Journal of Agricultural and Food Chemistry*. 2011;59:7902-7909.
91. Zhao, K. and Bartley, L.E. Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of *Arabidopsis*, poplar, rice, maize, and switchgrass. *BMC Plant Biology*. 2014;14:135.
92. Zhuang, Y. and Tripp, E.A. Genome-scale transcriptional study of hybrid effects and regulatory divergence in an F1 hybrid *Ruellia* (Wild Petunias: Acanthaceae) and its parents. *BMC Plant Biology*. 2017;17:15.

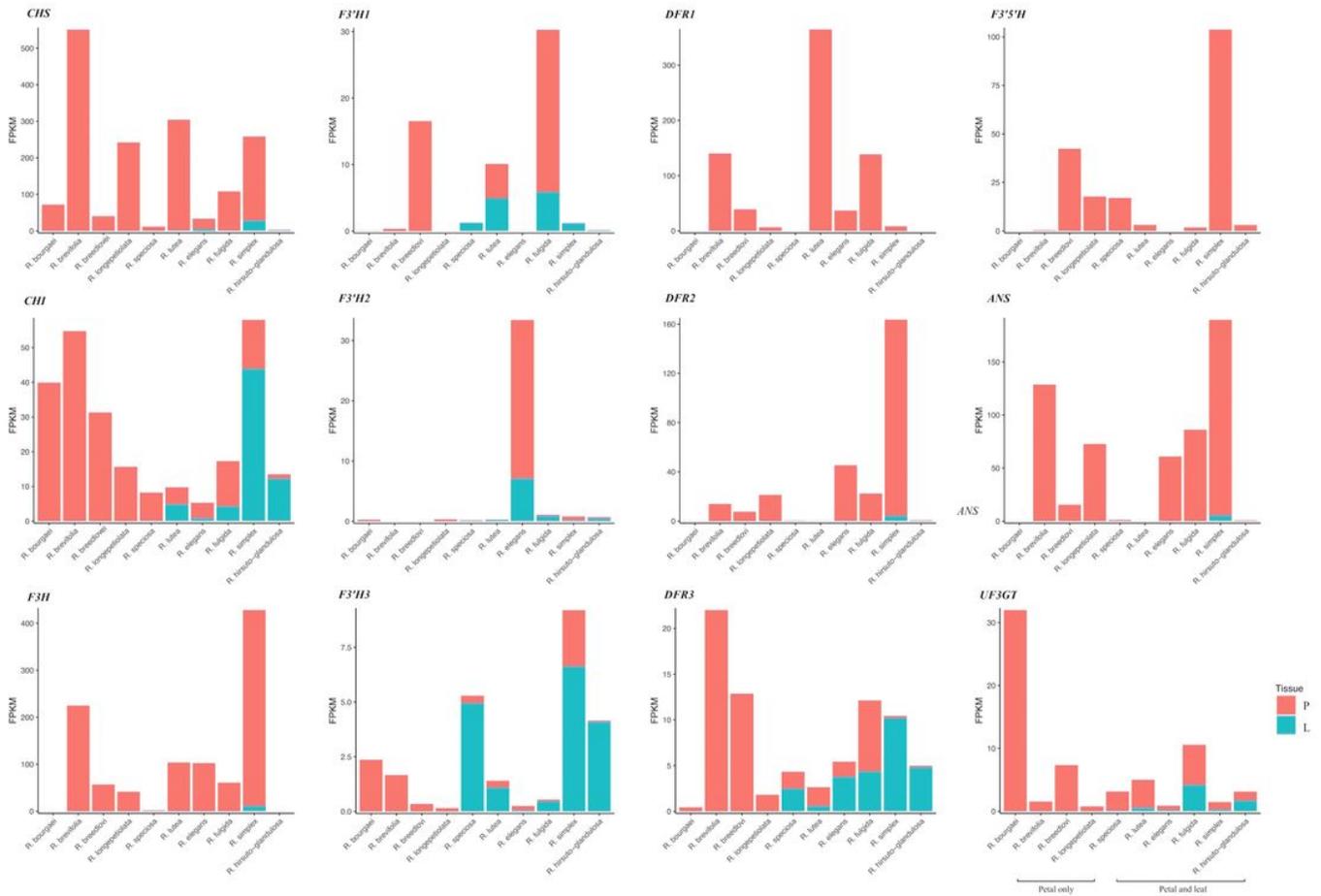
## Figures



**Figure 1**

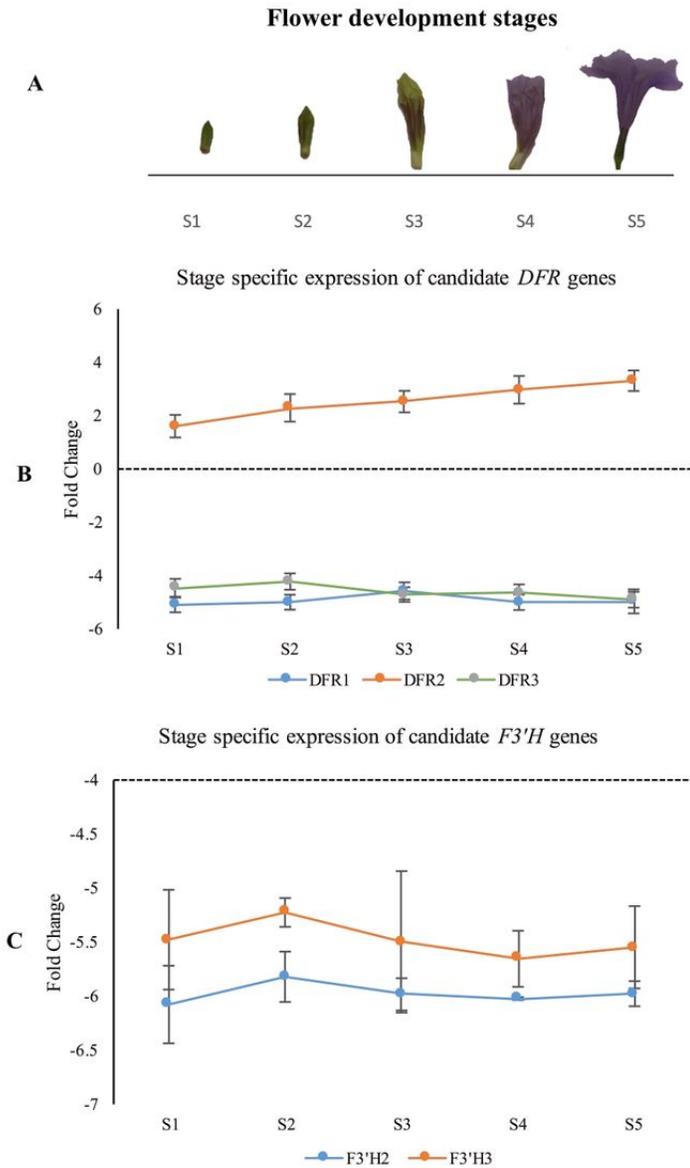
Phylogenetic tree depicting evolutionary relationships among 10 sampled species of *Ruellia* and principle component analysis (PCA) of transcriptome data from these species. A. Maximum likelihood tree was inferred using a hybrid transcriptome-RADloci approach newly implemented in this study. Briefly, raw, unassembled transcriptome reads were mined for the subset of reads that contained the GAATTC EcoRI overhang; these yielded a total 110,263 biallelic sites following data processing in pyRAD.

The ML tree was inferred using a GTRGAMMA model of sequence evolution implemented. Shown to the right of taxon labels are: photographs of flowers of the 10 study species as well as a shorthand coding of the major anthocyanin pigments present in petal tissues of these species (C=cyanidin + derivatives; D=delphinin + derivatives; P=pelargonidins+ derivatives). B. Gene expression profiles of ten species projected onto the first two principal components, showing two clear clusters of leaf (triangle) vs. corolla (circle) transcripts as well as two subclusters of corolla transcripts separated by PC2 that correspond to Clade I vs. Clade II in Figure 1A.



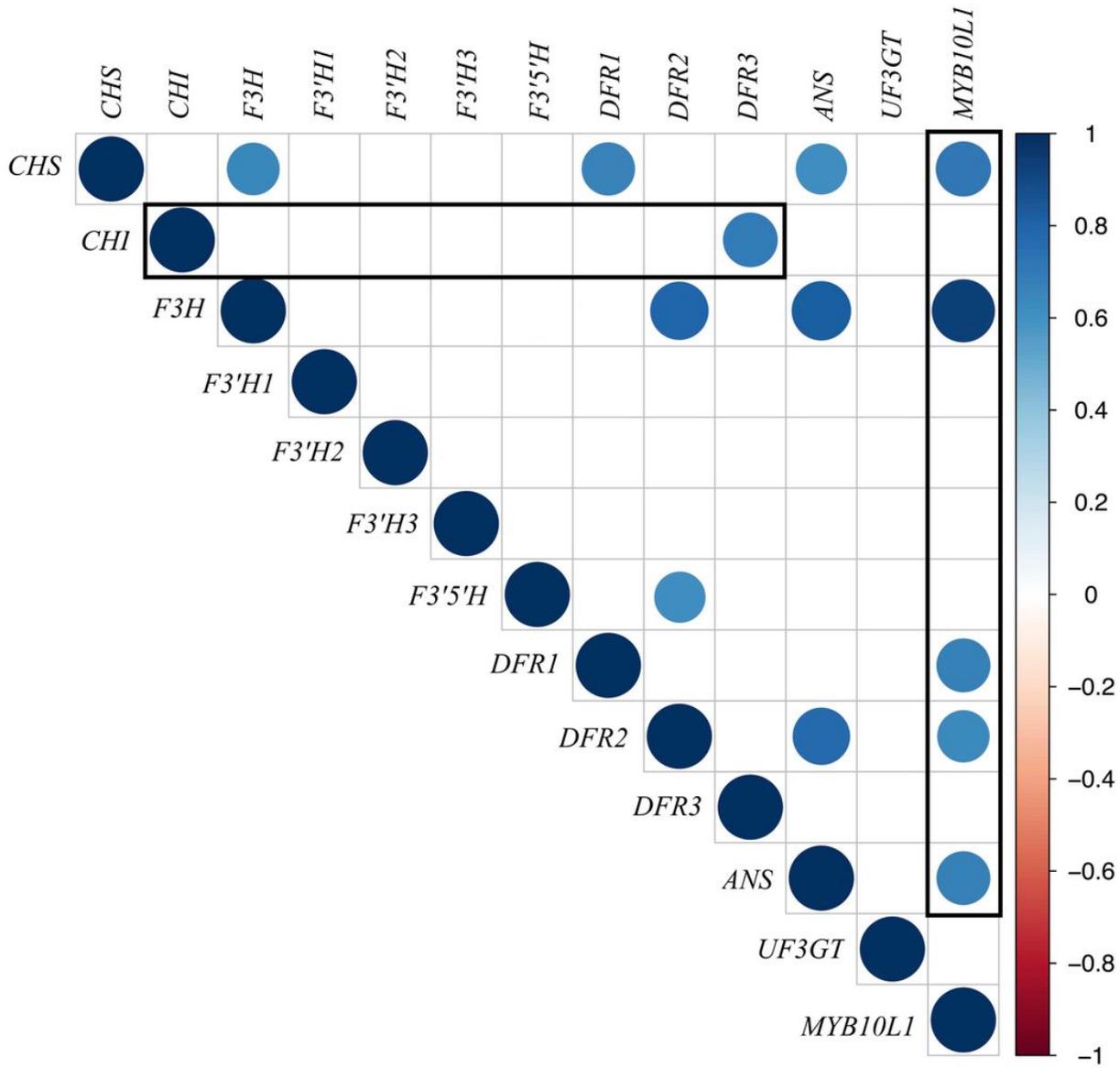
**Figure 2**

Transcript abundance of candidate anthocyanin structural genes identified in this study; both corolla and leaf transcriptome data were sequenced for six of the ten species of *R. speciosa*, *R. elegans*, *R. fulgida*, *R. simplex* and *R. hirsute-glandulosa*. Shown are petal (red) and leaf (cyan) transcript levels for several genes found as a single copy in among all samples (CHS, CHI, F3H, F3'5'H, ANS, UF3GT) as well as two genes found with three copies each (F3'H [copies 1–3] and DFR [copies 1–3]).



**Figure 3**

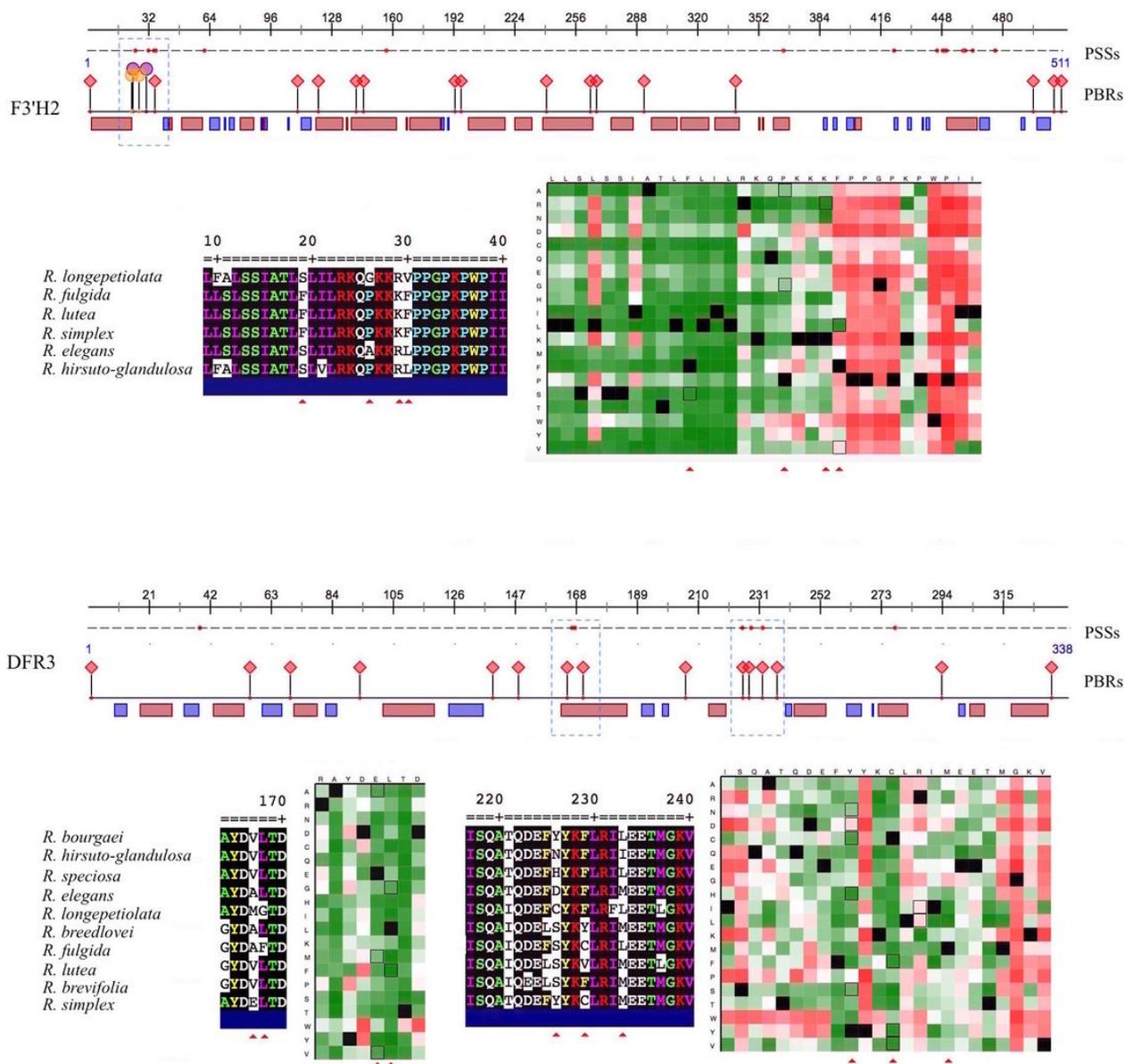
Relative expression of F3'H and DFR copies in *Ruellia simplex* during flower development based on qRT-PCR. Relative expression levels were calculated based on transcript abundance of the 18S gene in *Ruellia* using formula:  $Ct(18s) - Ct(\text{gene})$ . A. Five stages of floral development, from bud to fully anthetic flower. B. Stage-specific expression of DFR. C. Stage-specific expression of F3'H (note that F3'H1 was not expressed in *R. simplex*). Results show that DFR3 as well as F3'H2 and F3'H3 peak in expression levels during stage 2, suggesting possible coordinated regulation of these three loci.



**Figure 4**

A. Correlations among 13 conserved ABP-associated loci using the corrplot package in R ( $r \geq 0.65$ ). Predicted regulation blocks were highlight.





**Figure 6**

Detection of positive selection on several amino acid sites in F3'H2 and DFR3 in *Ruellia* and predicted effects on protein function. Upper Panels (both genes). Amino acid alignment positions shown at top. Black, dashed line with red dots denote sites under positive selection based on analyses implementing a M1 + M2 and M7 + M8 model of evolution. Protein binding regions (red diamonds) containing sites under positive selection marked with blue frames. yellow/orange circles stand for predicted nucleotide binding region, and the pink and blue rectangles stand for helix and strand respectively. Lower Panels (both genes). Expanded, view of blue frames showing amino acid substitutions and sites under positive selection, these marked by red triangles at bottom of alignment. To the right, prediction of functional effects of sequence variants conducted using *R. simplex* as a reference. Sites under positive selection marked by red triangles and highlighted with a black outline. Dark red (scores > 50): strong signal for a potential functional effect; white (scores < 50 or > -50): weak signal for potential effect; green (scores < -50): neutral or no effect.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3.xlsx](#)
- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [FigureS11.jpg](#)
- [FigureS10.jpg](#)
- [FigureS9.jpg](#)
- [FigureS8.jpg](#)
- [FigureS7.jpg](#)
- [FigureS6.jpg](#)
- [FigureS5.jpg](#)
- [FigureS4.jpg](#)
- [FigureS3.jpg](#)
- [FigureS2.jpg](#)
- [FigureS1.jpg](#)