Running title: Novel R2R3 MYB controls red anthocyanin production in *Iochroma*
[54 characters]

Title: Flavonol-regulating MYB underlies the evolution of red flowers in *Iochroma* (Solanaceae)

Authors: Lucas C. Wheeler[1,2,†], Maximilian Larter[3], Stacey D. Smith[1,†]
[1]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA
[2]Department of Biology, University of South Carolina, Columbia, SC, USA
[3]National Research Institute for Agriculture, Food and Environment (INRAE), Bordeaux, France; maximilian.larter@gmail.com
[†]authors for correspondence: lwheeler9@gmail.com, stacey.d.smith@colorado.edu

Date of submission: 11 March 2025
One table plus four figures
Word count (start of the introduction to the start of the acknowledgements, excluding materials and methods): 4149
Supplementary data: Six tables and eight figures

Highlight: A class of R2R3 MYB transcription factors previously known for their role in pepper peel antioxidants emerges as a key player in flower color evolution in a closely related genus.
[30 words]

Abstract

Anthocyanins, the pigments that give rise to blue, purple, red and pink colors in many flowers and fruits, are produced by the deeply conserved flavonoid biosynthesis pathway. The regulation of this pathway is thus fundamental for species differences in color across flowering plants, and a growing body of evidence implicates MYB transcription factors as key players activating or suppressing the production of different pigments. Here we demonstrate that a lineage of R2R3 MYBs that is closely related to well-known flavonol regulators (MYB12 members in subgroup 7) is the primary determinant of the shift from blue to red flowers in the genus *Iochroma*. Similar to its ortholog in *Capsicum*, this *Iochroma MYB12-like* gene controls the expression of flavonoid-3'-hydroxylase, the pathway branch point between red and blue pigments, and when down-regulated, results in redirection of flux toward red pigments. These results underscore the importance of transcription factor evolution in generating phenotypic novelty as well as the competitive nature of interactions among flavonoid pathway branches. In addition, our study demonstrates the effectiveness of RNAseq of segregating populations, in combination with other lines of evidence, for identifying novel functional variation. [186 words]

Keywords:

Transcriptomics, flavonoid biosynthesis, pigmentation, flower color, pelargonidin, gene regulation

Abbreviations:

DE: differentially-expressed; MYB: myeloblastosis; TPM: transcripts per million

Introduction

Phenotypic differences between species are often controlled by differences in the timing and patterns of gene expression (Kimura *et al.*, 2008, Des Marais and Rausher, 2010, Byers *et al.*, 2014). These differences in gene expression can arise through a variety of mechanisms, including changes in the cis-regulatory regions controlling expression (i.e., promoters, enhancers), changes in the expression or function of transcription factors, or post-transcriptional regulation (e.g., gene silencing). Many authors have argued that the cis-regulatory mutations will be favored during evolutionary transitions due to their modular architecture, allowing for altered expression in one context without pleiotropic effects in other contexts (Wray, 2007, Prud'homme *et al.*, 2006). However, functional changes in transcription factors can have similarly narrow consequences, depending on their specificity in terms of target genes and spatio-temporal patterns of expression (Lynch and Wagner, 2008, Panchy *et al.*, 2016, Auge *et al.*, 2019).

Plant MYB transcription factors comprise a prime example of a large and diverse gene family with highly specialized functions. Whereas animal and fungal genomes house at most a few dozen MYB genes, plant genomes contain hundreds of MYBs, even in diploid species (Shiu *et al.*, 2005, Feller *et al.*, 2011, Gates *et al.*, 2016). This expansion of MYB copies in plants is coupled with a diversification of functional roles, from defense, to coloration, to morphology (Ramsay and Glover, 2005, Wu *et al.*, 2022). Closely related MYBs often share similar regulatory functions, e.g., as activators or repressors of particular sets of target genes, but vary in their expression patterns, resulting in similar phenotypic effects albeit in different tissues or developmental stages (e.g., Millar and Gubler, 2005, Stracke *et al.*, 2007). Nevertheless, with the multitude of MYBs in every plant genome, new functional roles and patterns of diversification are continuing to be discovered (Sagawa *et al.*, 2016, Gates *et al.*, 2018, Mu *et al.*, 2024).

Among the subgroups of plant MYB transcription factors, those regulating floral coloration through the production of flavonoid pigments are among the best studied. The primary MYB activators of flavonoid synthesis fall into two subgroups of R2R3 MYBs: subgroup 7 (SG7) genes that regulate the ‘early’ genes of the pathway (e.g., CHS, F3H) and the branches leading to flavonol production (FLS), and the subgroup 6 (SG6) genes that regulate the ‘late’ steps of the pathway (e.g., DFR, ANS) leading to anthocyanin pigments (Dubos *et al.*, 2010, Albert *et al.*, 2014) (Fig. 1) Anthocyanins give rise to the red, purple and blue floral hues, while flavonols can modify these colors as co-pigments and provide UV-absorbing patterns, such as

nectar guides and bullseyes (Sheehan *et al.*, 2016, Todesco *et al.*, 2022). Thus, both types of compounds (anthocyanins and flavonols) are important contributors to floral coloration and are often jointly produced in developing petals.

While this general early/late regulatory architecture is well-conserved across flowering plants (Mol *et al.*, 1998, Schwinn *et al.*, 2014), the factors determining the type of flavonol or anthocyanin produced appear more variable across species, and perhaps for that reason, are not as well understood. Both flavonols and anthocyanins are produced at three hydroxylation levels (mono-, di-, and tri-) that have different spectral properties, and their relative expression depends on the expression of the so-called branching enzymes, F3'H and F3'5'H (Fig. 1). For example, when both enzymes are highly expressed, flowers will produce the tri-hydroxylated flavonoids, such as the blue delphinidin pigments, whereas when these enzymes are not present, flowers will produce the red pelargonidin pigments (Wessinger and Rausher, 2012; Fig. 1). The F3'H enzyme, which is responsible for conversion of DHK (the precursor of the flavonol kaempferol and the red pigment pelargonidin) into DHQ (the precursor of the flavonol quercetin and the purple pigment cyanidin), appears to be regulated by subgroup 7 MYBs in *Arabidopsis* and *Capsicum* (Stracke *et al.*, 2007, Wu *et al.*, 2023) and subgroup 6 MYBs in petunia and *Antirrhinum* (Albert *et al.*, 2011, Schwinn *et al.*, 2006). The other branching enzyme, F3'5'H, has been lost in many flowering plant lineages (e.g. morning glories, mustards) (Rausher, 2006, Falginella *et al.*, 2010), but in those which have retained the encoding gene, its expression is typically co-regulated with the late genes by the subgroup 6 MYBs (Albert *et al.*, 2011).

In the present study, we investigate the regulatory control of *F3'h* expression in *Iochroma* (nightshade family, Solanaceae), one of several genera in which red pelargonidin-producing flowers have evolved from blue delphinidin-producing ancestors. Previous work demonstrated that this flower color transition involved three genetic changes, including the down-regulation of F3'H, the evolution of substrate specificity in DFR, and the loss of the *F3'5'h* gene in the red-flowered species (Smith and Rausher, 2011, Smith *et al.*, 2013). Among these changes, the loss of *F3'h* expression has the largest effect on pigment production because it largely eliminates flux away from DHK, allowing anthocyanin production to be redirected towards pelargonidin (Smith and Rausher, 2011). Moreover, this shift in *F3'h* expression is due to a *trans*-regulatory mutation, as the genotype at the *F3'h* locus itself does not predict flower color in segregating populations (Smith and Rausher, 2011). This unknown regulator of *F3'h*, which segregates as a single gene, was termed the '*T*-locus' (Smith and Rausher, 2011).

Here we use a suite of genomic, transcriptomic, and biochemical approaches to identify candidates for the *T*-locus responsible for the shift toward pelargonidin production and in turn,

the evolution of red flowers in *Iochroma*. Using biochemical and expression data, we first sorted individuals from a backcross population by pigment phenotype and corresponding difference in *F3'h* expression. Next, we searched the floral transcriptomes of these two pools of individuals for genes that match the predicted allelic pattern (e.g., homozygous for the red-flowered parent allele in the pink/red-flowered pool) and show the predicted association with *F3'h* expression. Our analyses point to a single R2R3 MYB transcription factor that is related to the MYB12 members of Solanaceae subgroup 7 MYBs but falls in a deeply diverged clade, only functionally characterized in chili peppers. As we discuss, these results suggest that the subgroup 7 MYBs may be much more diverse than previously known and play an underappreciated role in flower color evolution through their effects on flavonol production.

Materials and methods

Source populations and phenotyping

Individuals of the blue-flowered *I. cyaneum* were crossed with the red-flowered *I. gesnerioides* to create segregating populations to dissect the genetic basis of their flower color differences (Smith and Rausher, 2011). The blue-flowered state is ancestral in *Iochroma* and corresponds to the production of delphinidin-derived anthocyanins, while the red-flowered derived state involves the production of pelargonidin-derived anthocyanins (Fig. 1; Smith and Rausher, 2011). The *I. cyaneum* parent was grown from seed from a cultivated accession from the Missouri Botanical Gardens, originally collected by W. G. D'Arcy, and the *I. gesnerioide*s parent was grown from the Solanaceae Germplasm collection in the Botanical Garden of Nijmegen (accession number 944750129). Herbarium vouchers for these accessions are Smith 265 and 266 (WIS), respectively. A single F1 was backcrossed to the *I. gesnerioides* parent, and progeny from the resulting backcross population were genotyped at *F3'5'h* and *Dfr* (Smith and Rausher, 2011; Table 1). Anthocyanin production was previously characterized using HPLC and revealed three pigment phenotypes (purple-flowered individuals producing primarily cyanidin, pink-flowered individuals producing mostly pelargonidin, and red-flowered individuals producing almost entirely pelargonidin) (Smith and Rausher 2011). The purple-flowered individuals share high *F3'h* expression and are inferred to carry a dominant 'blue' allele at a segregating trans-acting factor (the '*T*-locus', Smith and Rausher 2011) (Table 1).

Biochemical phenotyping and RNA-seq of backcross individuals

We performed RNA-Seq on corolla tissue from 24 backcross individuals segregating for the putative *T*-locus. We sampled 12 individuals with each inferred *T*-locus genotype: *Tt* corresponding to one dominant 'blue' allele and high *F3'h* expression or *tt* corresponding to two recessive red alleles and low *F3'h* expression (Table 1). We divided these 12 among the possible genotypes at the other two loci that affect flower color in this cross (*Dfr* and *F3'5'h*). *Dfr* shows functional specialization, with the red allele specialized for activity on DHK (Smith *et al.* 2013), while *F3'5'h* is absent from the red parent genome (Smith and Rausher, 2011). With four possible combinations at these other two loci (*Dd*/*F*-, *Dd*/–, *dd*/F-, *dd*/–), we sampled three biological replicates of each within the groups of 12 (Table 1). We included all possible genotypic combinations at the three loci influencing flower color in order to isolate the *T*-locus while balancing across the effects of these other loci. For RNA extraction, we flash-froze corolla tissue from buds of roughly 1.25cm in length, which is equivalent to Petunia bud Stage 5 (Pollak *et al.*, 1993). This developmental stage shows expression of both early and late pathway genes in the anthocyanin pathway (Larter *et al.*, 2018). Total RNA was extracted with the Spectrum Total RNA extraction kit (Sigma, St Louis, MO). Library preparation and 150-base-pair paired-end mRNA sequencing was carried out by Novogene (Sacramento, CA).

Identifying SNPs associated with flower color and F3'h expression

We used the reference genome assembly for *Iochroma cyaneum* (Powell *et al.*, 2022) to call SNP variants and filter the RNASeq dataset for candidates for the *T*-locus. RNAseq reads were aligned with STAR (Dobin *et al.*, 2013), and the resulting BAM files were used as input for bcftools mpileup with default settings to call variants. We filtered variants by base call quality, only retaining variants with quality score greater than or equal to 20. We used the resulting VCF file for subsequent analyses of associations with the color phenotype.

We first split the filtered VCF files into two subsets, one for all samples with purple cyanidin-producing flowers (inferred *Tt* genotype at *T*-locus) and one for those with pink or red mostly pelargonidin-producing flowers (inferred *tt* genotype at *T*-locus) (Table 1). In order to identify SNPs that differ between these two pools, we used *pyvcf* (Casbon, 2012) to filter the variants to include only those that are present in all "*Tt*" individuals and not present in any "*tt*"

individuals. This strict criterion resulted in a set of SNPs that perfectly co-segregate with the high or low *F3'h* expression (see Results). Most of the SNPs are located on chromosome 5, but some mapped to smaller scaffolds that were not incorporated into the reference assembly (Supplementary Fig. S1). We then used promer from Mummer4 (Marçais *et al.*, 2018) and D-genies (Cabanettes and Klopp, 2018) to align these scaffolds back to the *I. cyaneum* and tomato reference genomes.

In addition to this filtering approach, we performed a case-control GWAS with the variant calls in GEMMA (Zhou and Stephens, 2012). We set phenotypes to 0 (purple-flowered *Tt* plants) or 1 (pink/red-flowered *tt* plants) and fit a univariate linear mixed model with the full set of variants. We then plotted the location of all analyzed variants on the assembled *I. cyaneum* chromosomes and identified variants with significant phenotypic associations.

Co-expression of candidate genes with *F3'h*

We predicted that if the *T*-locus is a transcriptional regulator, its expression will likely track that of *F3'h* in the segregating backcross. Thus, we used expression data from the 24 transcriptomes to quantify levels of expression and test for correlations between *F3'h* and loci carrying associated SNPs (previous section). We first created a *de novo* transcriptome for the blue-flowered parent (*I. cyaneum*) to ensure that we captured all expressed genes. For this assembly, we used single-end Illumina RNA-seq data from reproductive, seed, and vegetative tissues from *I. cyaneum* from a previous study (Powell *et al.*, 2022) and assembled the transcripts using the pipeline developed in Wheeler et al (2022). Briefly, we corrected read errors in the 128,433,717 raw reads using Rcorrector (Song and Florea, 2015) and removed unfixable reads using *unfixable_filter.py* (Yang and Smith, 2014). We trimmed adaptor sequences from the filtered reads using Trimmomatic (Bolger *et al.*, 2014) and used the trimmed reads for de novo assembly with Trinity (Grabherr *et al.*, 2011). We removed apparent chimeric sequences using *run_chimera_detection.py* (Morales-Briones *et al.*, 2021), with a reference BLAST database consisting of sequences from *Arabidopsis*, *Solanum*, and *Petunia*. We then used Corset (Davidson and Oshlack, 2014) to cluster transcripts and *filter_corset_output.py* (Yang and Smith, 2014) to remove redundant transcripts. Finally, we predicted complete CDS from the Corset-filtered transcripts using TransDecoder (Haas *et al.*, 2013).

Next, we quantified gene expression by pseudo-aligning reads from each backcross individual to the predicted CDS in the transcriptome using Salmon (Patro *et al.*, 2017). We

calculated estimated read counts and TPM for each transcript. We imported Salmon quant files, partitioned by inferred *T*-locus genotype (*Tt*/*tt*), into DEseq2 with *tximport* (Soneson *et al.*, 2015) and used the *DESeqDataSetFromTximport* function to create a DEseq analysis object, with treatments corresponding to the *T*-locus genotype. We quantified differential expression between these subsets using the *DESeq* function. We filtered the resulting transcripts by adjusted p-value with a significance threshold of padj=0.05 to identify significant DE transcripts.

We used WGCNA (Langfelder and Horvath, 2008) to identify modules of co-expressed genes, as we predict that the *T*-locus would be co-expressed with *F3'h* and possibly other flavonoid biosynthesis genes. WGCNA computes pairwise correlation coefficients, which then are converted to an adjacency matrix with the raw values raised to a soft-thresholding power (β) to approximate a scale-free network. For our data, we selected a β of 7, which corresponds to an R^2 value of 0.88 with the scale-free model and a mean connectivity of 20.4 (Supplementary Fig. S2). We initially used blockwise module detection on the full de novo transcriptomic dataset of 19,184 genes, and from this first pass, we retained modules with a correlation of 0.2 or greater with the trait of interest (color phenotype/inferred *T*-locus genotype). The filtered dataset contained 4854 genes, which allowed us to examine smaller modules (minimum size of 20 genes). After hierarchical clustering, we merged modules that were 90% similar and re-calculated correlations between the module eigengenes and the trait.

We exported the resulting module containing *F3'h* to Cytoscape format using *exportNetworkToCytoscape* and extracted the topology overlap matrix (TOM) edge weights. We plotted the distribution of weights for edges containing *F3'h* and for all other edges and used Z-scores to capture how extreme each co-expression relationship is within the context of the module. We considered genes that emerged from the association mapping (above) and presented significantly correlated expression with *F3'h* as strong candidates for the *T*-locus.

Phylogenetic analysis of *MYB12-like* genes and other SG7 MYBs

Our combined analyses of SNP association and gene expression strongly implicated an R2R3 MYB, which we refer to as *Iochroma cyaneum MYB12-like* following the nomenclature in *Capsicum* (see Results). As R2R3 MYBs comprise a large group of functionally distinct transcription factors, we carried out phylogenetic analysis to identify the most closely related copies in other model Solanaceae and in *Arabidopsis*. We used BLAST searches to retrieve the top hits from tomato, potato, groundcherry, chilipepper*, Nicotiana benthamiana*, and *Arabidopsis thaliana* and created a protein alignment with MAFFT v. 7 (Katoh and Standley,

2013) using default settings. As BLAST results suggested that the most similar sequences belonged to the flavonoid-regulating subgroup 7 (SG7) MYBs, we included the R2 and R3 MYB domains through to the SG7 motif (Stracke *et al.*, 2007, Stracke *et al.*, 2001) in the alignment. The downstream positions were trimmed as they were hypervariable and could not be confidently aligned. We estimated a maximum-likelihood phylogeny using this SG7 amino acid alignment with the best-fitting model of amino acid substitutions (Q.plant+G4) and 1000 bootstrap replicates in IQ-TREE 2.3.6 (Nguyen *et al.*, 2015, Minh *et al.*, 2020). We rooted the resulting topology on the lineage leading to the clade containing AtMYB111, AtMYB11 and AtMYB12 (Schilbert and Glover, 2022).

Based on this broader phylogenetic analysis, we identified a set of MYBs most closely related to the candidate locus. We next estimated a tree from full length coding sequences (CDS) from those closely related copies, which are more easily aligned. We included six additional Iochrominae sequences assembled by mapping reads from floral bud transcriptome data onto the *Iochroma cyaneum* genome sequence, again using STAR. Each of these six species is represented by two biological replicates; a consensus of the two was used for the phylogeny and the replicates were used to estimate the levels of MYB12-like expression in each species using Salmon as above. We estimated the maximum likelihood tree from the CDS alignment with IQ-TREE, using the best-fitting model of nucleotide substitutions (TIM3+F+G4).

Statistical analysis

We conducted statistical analyses using R version 4.2.3 and additional software as described above. All scripts along with input and output files are included in a single OSF repository to allow the results to be easily replicated (https://osf.io/j5m8f/). Raw RNASeq data from the 24 backcross individuals (Table 1) as well as from the six Iochrominae species is available on NCBI's short read archive as BioProjects PRJNA1092111 and PRJNA1102413, respectively.

Results

Localization of associated SNPs with flower color in the *Iochroma* genome

We recovered 92 SNPs that perfectly co-segregate with the two phenotypic pools, i.e., distinguish purple-flowered *Tt* and pink/red-flowered *tt* pools). The majority of these SNPs (49,

53%) fall on chromosome 5 of the *I. cyaneum* reference assembly. We also found 28 SNPs mapping to a roughly 620 Kb scaffold (00085) and the remainder (15) mapping to three additional unincorporated scaffolds (Supplementary Fig. S1). Our subsequent analyses suggest that these scaffolds represent segments of chromosome 5 that were not included during the assembly process (Powell *et al.*, 2022). For example, scaffold00085 aligns well with tomato chromosome 5 (Supplementary Fig. S3), and 95% of the CDS retrieved from that scaffold have top hits on tomato chromosome 5. This region appears nested within the larger region of *I. cyaneum* chromosome 5 where most of the SNP associations are clustered (Fig. 2). The three smaller scaffolds with associated SNPs (Supplementary Fig. S1) also BLAST to tomato chromosome 5 and were also likely excluded during assembly. Thus, all SNPs recovered from the co-segregation analysis appear to be localized along a small region of *I. cyaneum* chromosome 5.

We carried out a case-control GWAS using the same set of variant calls. This analysis similarly retrieved associations exclusively on chromosome 5, with significant hits in the gene-dense region in the last 500kb of the chromosome (Fig. 2; Supplementary Fig. S4). This region of the genome contains 468 gene models (Supplementary Table S1), 352 of which are functionally annotated in the genome (Supplementary Table S2). Twenty-eight of these genes are annotated as transcription factors and only one corresponds to a known group of flavonoid regulators. This locus (IC05g034110) is annotated as a MYB111 transcription factor based on similarity with AtMYB111, a flavonol-regulating subgroup 7 MYB (Stracke *et al.*, 2007); we will refer to this gene as *Iochroma cyaneum MYB12-like* (*IcMYB12-like*) based on the phylogenetic analysis (see below). The region also contains copies of one of the upstream pathway enzymes, chalcone synthase (CHS), as well as UDP-glycosyltransferase (UGT), which can glycosylate various flavonoids.

Patterns of differential expression and co-expression

Our DEseq2 analysis identified 58 significantly differentially expressed transcripts between the two phenotypic pools in our backcross (Supplementary Table S3). The MYB transcription factor *IcMYB12-like* appears as the sixth most strongly differentially expressed gene between the pools (log_2-fold change = -6.35, or ca. 82-fold lower expression in the pink/red pool). Its putative target, *F3'h*, is the eighth most differentially expressed (log_2-fold change = -5.83, or ca. 57-fold lower expression in the pink/red pool). Note that the expression of *F3'h* in many of these individuals was previously measured with qPCR (Table 1; Smith and Rausher, 2011); this

analysis confirms the strength and degree of the differential expression between individuals presenting the alternate pigment phenotypes (Fig. 3A). A similar degree of differential expression was found for *Fls* between the two pools, and three other flavonoid pathway genes (*Chs*, *UGt*, and 3Gt) also appear among the list of significantly DE genes (Supplementary Table S3). These patterns could indicate some degree of regulatory control of *IcMYB12-like* over other pathway steps.

In addition to examining DE genes between the two phenotypic pools, we explored co-expression of genes across the entire set of 24 backcross individuals. If *IcMYB12-like* indeed activates floral *F3'h* expression, we expect the two genes to show correlated expression and to belong to the same co-expressed module of genes. Consistent with this prediction, our WGCNA analysis recovered a module of 52 genes containing *F3'h* and *IcMYB12-like* (Supplementary Table S4). Out of the 34 modules found in the analysis, the module is the only one significantly correlated with the pigment phenotype (purple vs. pink/red, R^2=-0.92, p=$1e^{-10}$, Supplementary Fig. S5). Within this module, *IcMYB12-like* is tightly co-expressed with *F3'h* (Fig. 3B, Supplementary Table. S4). The connectivity between *F3'h* and *IcMYB12-like*, measured as topological overlap matrix (TOM) values from the WGCNA analysis, was the second highest in the set of all edges involving *F3'h* (Z-score: 1.96) with only the edge connecting *F3'h* and *Fls* having a higher value (Z-score: 2.56) (Supplementary Fig S6). We also found a tight connection between *IcMYB12-like* and *Fls* (Z-score: 1.83), suggesting that both *F3'h* and *Fls* are both regulated by *IcMYB12-like.* Four other flavonoid pathway genes, *Chs*, *Chi*, *3Gt*, and *Ugt* appear in the module associated with the phenotype, and all except for *Chi* are directly connected to *IcMYB12-like* (Supplementary Table S4; Fig. 3B). Eight other loci within phenotype-associated module are connected to *IcMYB12-like* (*e.g.*, DETOX-35-like-2 and the F-box protein At5g07610-like), although no functional connection is known. Three of the genes connected to *IcMYB12-like* (R1A-10, the F-box protein At5g07610-like and UGT) fall in the same genomic region as *IcMYB12-like* (Supplementary Table S2), suggesting that these co-expression patterns may be related to co-localization within the genome (Michalak, 2008). Indeed, differentially expressed transcripts are clustered around *IcMYB12-like* (Supplementary Fig. S7). Nevertheless, both *F3'h* and *Fls* occur outside of the region containing *IcMYB12-like* (Fig. 2), excluding co-localization as an explanation for their strong co-expression with *IcMYB12-like.*

Phylogenetic relationship of *IcMYB12-like* to other MYB transcription factors

We used BLAST searches to retrieve similar sequences to *IcMYB12-like*. The top hits from Arabidopsis and Solanaceae genomes corresponded to members of subgroup 7 of R2R3 MYB transcription factors (Stracke *et al.*, 2001). This subgroup controls flavonol production in Arabidopsis (Stracke *et al.*, 2007) by regulating upstream steps such as CHS, CHI, and FLS, as well as the glycosyltransferases that stabilize these products. Maximum-likelihood analysis revealed that *IcMYB12-like* and highly similar sequences from pepper and potato are closely related to subgroup 7 but fall in a separate subclade, with strong support (Fig. 4A). Subgroup 7 MYBs have been well characterized in Solanaceae (e.g., Ballester *et al.*, 2010, Song *et al.*, 2019) and appear functionally similar to their orthologs in *Arabidopsis*. *Iochroma* possesses an SG7 MYB that is closely related to these well-characterized MYB12 genes (IC05g030210, Fig. 4A) in addition to the divergent sequence (IC05g034110), which we refer to as a *MYB12-like* gene following the naming of *CaMYB12-like* in *Capsicum* (CA05g18430 in Fig. 4A).

Using additional BLAST searches beyond nightshade crops, we identified an additional member of the *MYB12-like* clade in *Lycium*, which we used to root the phylogeny including the additional Iochrominae sequences (Fig. 4B). The topology is similar to the species tree (Deanna *et al.*, 2019) although most of the branches are unsupported given that the sequences present few differences (Supplementary Fig. S8). Examining the expression of *MYB12-like* in these taxa in relation to their floral flavonol production (Larter *et al.*, 2019), we observed that species with higher *MYB12-like* expression also produce higher amounts of flavonols (Fig. 4B), which are mainly quercetin glycosides (Berardi *et al.*, 2016). This pattern aligns well with the proposed function of *MYB12-like* in activating *F3'h*, which in turn produces DHQ, the precursor of quercetin (Fig. 1)

Discussion

This study aimed to identify the gene underlying the so-called *T*-locus, which acts as a transcriptional regulator of *F3'h* to determine flower color in the nightshade genus *Iochroma*. By carrying out RNASeq of floral bud tissue from multiple backcross individuals with *T*-locus genotypes assigned based on flower color (Fig. 1B), we pinpointed an R2R3 MYB transcription factor as the strongest candidate for the *T*-locus. First, our SNP-association studies narrowed the candidate region to 10Mb near the end of chromosome 5 (Fig. 2; Supplementary Fig. S4). This region of the genome contains 468 gene models, including 28 annotated as transcription factors. Among these, only one of these corresponds to a class of genes, SG7 MYBs, known to be involved in regulating flavonoid biosynthesis. This *MYB12-like* gene shows tightly correlated

expression with *F3'h* across the backcross (r=0.91, Fig. 3A). Indeed, these two genes emerge as part of a compact module in transcriptome-wide co-expression analyses, with *F3'h* having a stronger connection with *MYB12-like* than any other gene in the floral transcriptome with the exception of *Fls* (Fig. 3B; Supplementary Fig. S6). Given that the effect of the *T*-locus could be due to coding sequences changes only, this set of analyses cannot conclusively eliminate other candidate transcription factors in the associated region of the genome. Nevertheless, our phylogenetic analyses identify *IcMYB12-like* as an ortholog of chillipepper *CaMYB12-like*, a recently characterized flavonoid regulator, which like its *Iochroma* ortholog, acts as a positive regulator of *F3'h* (Wu *et al.*, 2023). Together, these lines of evidence argue that the *T*-locus corresponds to the *MYB12-like* gene in *Iochroma*, which drives the origin of red flowers by altering floral flavonoid composition. Below we discuss how these findings contribute to our broader understanding of flower color evolution.

The role of MYB transcription factors in shaping floral hue

While genetic studies of flower color have long implicated subgroup 6 R2R3 MYBs as the major determinants of floral pigment intensity (e.g., Quattrocchio *et al.*, 1999, Schwinn *et al.*, 2006, Streisfeld *et al.*, 2013), work on the genetic basis of changes in floral hue has implicated a wide variety of molecular mechanisms (Wessinger and Rausher, 2012, Berardi *et al.*, 2021, Quattrocchio *et al.*, 2006). Differences in the type of anthocyanins produced, which in turn influence the type of flower color, can arise from shifts in gene regulation (either in *cis*- or *trans*-) as well as changes in the function of pathway enzymes (Hopkins and Rausher, 2011, Wessinger and Rausher, 2015, Smith and Rausher, 2011, Smith *et al.*, 2013, Wheeler *et al.*, 2023). Nevertheless, the identity of transcription factors that influence the type of anthocyanin produced (as opposed to the overall amount) has remained nebulous.

Because of the shared precursors within the flavonoid pathway, subgroup 7 (SG7) transcriptional regulators of flavonol production can directly influence anthocyanin production, and, as shown in the present study, the type of anthocyanin produced as well. The deeply conserved structure of the pathway presents multiple branching points where a single precursor can be converted in different products depending on the enzymes present and their properties (Tohge *et al.*, 2013, Winkel-Shirley, 2001). The colorful anthocyanins share dihydroflavonol precursors (DHK, DHQ, DHM) with flavonols, creating the potential for competition between DFR and FLS for these substrates (Fig. 1A). Thus, the upregulation of SG7 MYBs and, in turn, their targets (*Chs*, *Chi*, *F3h*, *F3'h*, *Fls* and sometimes *F3'h*) generally reduces anthocyanin

production in favor of flavonols to produce paler flowers (Holton *et al.*, 1993, Yuan *et al.*, 2016, Wheeler *et al.*, 2023). The precise effect of altering the expression of SG7 MYBs on flower color will, however, depend on their target genes and the substrate preferences of multifunctional pathway enzymes (e.g., DFR, FLS).

In the case of *Iochroma*, the ability of the SG7 *MYB12-like* gene to alter flower color is likely due to the combination of a narrowing of target genes and strong substrate preferences among downstream enzymes. While the pepper *CaMYB12-like* gene activates a broad suite of early genes (*Chs, Chi, F3h, F3'h, Fls, 3GT*, Wu *et al.*, 2023), the *Iochroma* ortholog only shows strong co-expression with *F3'h* and *Fls* (plus weaker co-expression with *Chs*, *3Gt* and, *Ugt*), indicating a reduced suite of targets. The broad upstream action by *CaMYB12-like* is similar to that of the other well-known SG7 MYBs in Solanaceae (*SlMYB12* in tomato, Ballester *et al.*, 2010, Fernandez-Moreno *et al.*, 2016; *NtMYB12* in tobacco, Song *et al.*, 2019), suggesting that coordinated regulation of 'early' genes represents the ancestral state and that the functional shift toward specificity has occurred along the *MYB12-like* lineage leading to *Iochroma*. Accordingly, the loss of *MYB12-like* expression in *I. gesnerioides* flowers is not associated with a complete disruption in floral flavonoid pigment production (Berardi *et al.*, 2021, Larter *et al.*, 2019), but a targeted reduction in DHQ through lower *F3'h* expression. The resulting accumulation of DHK is not converted to kaempferol, likely because of coordinated loss of *Fls* expression and its low preference for DHQ, at least in the berry-fruited Solanaceae like *Iochroma* (Bovy *et al.*, 2007, Berardi *et al.*, 2016, Rosa-Martínez *et al.*, 2023). Instead, this DHK precursor is converted to red pelargonidin pigments by the DFR enzyme, which in *I. gesnerioides*, is specialized for DHK (Smith *et al.*, 2013). Smith *et al.* (2013) hypothesized that, during the evolutionary transition from blue to red flowers, the *trans*-regulatory loss of *F3'h* expression occurred first, allowing the flux to shift toward red pigmentation. Under this scenario, the selection would be expected to favor increased activity of DFR on DHK to allow efficient conversion to red pelargonidins.

This *MYB12-like*-mediated biochemical trade of blue anthocyanins plus flavonols for red anthocyanins alone may have also carried ecological consequences for relationships with pollinators. In addition to acting as co-pigments, flavonols increase floral UV-absorbance, which is attractive to moth pollinators (Sheehan *et al.*, 2016), and can also enhance fly and bee visitation if associated with floral patterning (Koski and Ashman, 2014). Indeed, insects comprise only 10% of pollinator visits to *Iochroma gesnerioides* compared hummingbirds, which account for 90% of visits (Smith *et al.*, 2008). This lack of UV-absorbing flavonols is isolated to *I. gesnerioides* flowers as the leaves produce comparable amounts of flavonols (specifically

quercetin) as the blue-flowered *I. cyanuem* (Berardi *et al.*, 2016) and the expression of *F3'h* is actually higher in *I. gesnerioides* leaves than in those of *I. cyanuem* (Smith and Rausher, 2011). The targeted effects of *MYB12-like* on floral flavonols may have thus created an accessible evolutionary pathway to red flowers, given that a loss of quercetin across the entire plant would carry significant negative pleiotropic effects (Ryan *et al.*, 2001, Singh *et al.*, 2021).

MYB transcription factors in the evolution of species differences

Closely related species of flowering plants are often distinguished by subtle differences in their reproductive organs, e.g., in the color, shape, scent, or pubescence of flowers or fruits. MYB transcription factors control many of these aspects of morphological development and epidermal cell fate (Ramsay and Glover, 2005, Hileman, 2014), which may help to explain their prevalence in underlying fixed differences between species (e.g., Preston *et al.*, 2011, Castillejo *et al.*, 2020, Gates *et al.*, 2018, Yarahmadov *et al.*, 2020). In fact, MYB transcription factors may act as speciation genes when the phenotypic differences resulting from changes in their function or expression leads to reproductive isolation (Streisfeld *et al.*, 2013, Sheehan *et al.*, 2016, Lüthi *et al.*, 2022). Through its simultaneous effects on visible anthocyanins and UV-absorbing pigments, changes in floral *MYB12-like* expression could have played a role in species divergence, although the split between the red-flowered clade containing *I. gesnerioides* and its blue-flowered relatives likely occurred 5 to 10 million years ago (Huang *et al.*, 2023), and the two lineages no longer occur in hybrid zones. The *I. arborescens* complex (the "A" clade sensu Smith and Baum (2006)) presents a stronger opportunity for dissecting the role of *MYB12-like* in floral isolation as red-flowered, low-flavonol primarily-hummingbird-pollinated species (e.g., *I. edule*) co-occur and hybridize with flavonol-rich insect-pollinated species (e.g., *I. arborescens*) (Smith *et al.*, 2008; Fig. 4B).

Cis-regulatory mutations involving MYBs appear to be a major target for evolutionary transitions, and our results suggest that regulatory changes, as opposed to functional variation, drive the effects of *MYB12-like* on flower color in *Iochroma*. First, the *MYB12-like* sequence from *I. gesnerioides* shows 6 fixed amino acid differences from its closest blue-flowered relative (*I. calycinum*), however all but one of these variants (a threonine indel close to the 3' end, Supplementary Fig. 7) are segregating across *Iochroma* species with both low and high flavonol accumulation (Fig. 4B). Moreover, *MYB12-like* expression levels are strongly predictive of pathway activity. Within the backcross, the red *I. gesneroides* parent allele of *MYB12-like* is expressed at extremely low levels, which in homozygous state translates to a near absence of

F3'h expression (Fig. 3A). This relationship extends above the species level, where lower levels of *MYB12-like* expression appear to be associated with lower levels of quercetin flavonols (Fig. 4B). Beyond *Iochroma*, *cis*-regulatory mutations at MYB transcription factors frequently contribute to within and among species differences in flower color (Martins *et al.*, 2017, Streisfeld *et al.*, 2013, Fattorini and Ó'Maoiléidigh, 2022), a pattern that has been attributed to their comparatively limited pleiotropic effects (Sobel and Streisfeld, 2013). Nevertheless, the precise changes in the MYB promoters are unknown in these natural systems. Identifying the causal mutation(s) will require fine dissection of the promoter region along with *in vivo* or *in vitro* assays of various constructs (e.g., Espley *et al.*, 2009, Jia *et al.*, 2021). Although transformation remains challenging outside of model systems, pinpointing these causal variants is important for ultimately understanding how and why MYBs and the modules they control can be deployed in new developmental contexts.

Conclusions

As a powerful group of antioxidants, flavonols have long been the focus of efforts in plant breeding, resulting in a detailed understanding of the subgroup 7 MYBs that largely control their expression across flowering plants. Within the nightshades, the best known of these MYBs are the orthologs of MYB12, which contribute to stress tolerance in tobacco (Song *et al.*, 2019) and the color of the fruit peel in tomato (Ballester *et al.*, 2010) via their effects on flavonoid production. This gene family is also expressed in Solanaceae flowers (Zheng *et al.*, 2021), activating early branches of the pathway to provide both flavonol co-pigments and the substrates for anthocyanin biosynthesis. Our work reveals that, in addition to this canonical 'MYB12' group of SG7 MYBs, *Iochroma* flowers also express a more divergent 'MYB12-like' lineage that has evolved narrow specificity for FLS and F3'H. This specialization, together with flower-specific expression, allows *IcMYB12-like* to act as the switch between blue and red flowers. Piecing together the origin of this gene's role in floral flavonoid production will require additional sampling of closely related genomes as they emerge. Still, our study together with the larger body of literature underscores how the diversification of MYB transcription factors is intimately connected to the diversification of plant phenotypes, from crop varieties to interspecies differences.

Supplementary data

Table S1. Gene models in the genomic region containing phenotype-associated SNPs.
Table S2. Functionally-annotated genes in the phenotype-associated genomic region.
Table S3. Differentially expressed transcripts from the DESeq analysis.
Table S4. Nodes and edges within phenotype-associated module from WGCNA analysis.
Table S5. Gene expression for *MYB12-like* and floral flavonol content for six Iochrominae species.
Table S6. Gene names and descriptions for phylogenetic analysis.
Fig. S1. Unincorporated scaffolds containing phenotype-associated SNPs
Fig. S2. Scale independence and mean connectivity for soft threshold selection for WGCNA analysis.
Fig. S3. *Iochroma* chromosome 5 and scaffold00085 mapped to tomato chromosome 5.
Fig. S4. Manhattan plots for associations with flower color across all *I. cyaneum* chromosomes.
Fig. S5. Module-trait relationships from the WGCNA analysis.
Fig. S6. Distribution of connectivity (TOM) values between DE transcripts.
Fig. S7. Elevated differential expression near *MYB12-like* locus.
Fig. S8. Amino acid alignment for MYB12-like sequences from Iochrominae and other Solanaceae.

Acknowledgements

The authors thank Adrian Powell (Boyce Thompson Institute) for providing access to the *Iochroma cyaneum* genome through the SolGenomics website.

Author contribution

SDS, LCW: conceptualization, methodology; SDS, LCW, ML: data collection; SDS, LCW: formal analysis; SDS: writing – original draft; SDS, LCW: writing - review & editing; LCW: data curation; SDS: funding acquisition.

Conflict of Interest

No conflict of interest declared.

Funding Statement

This work was supported by the National Science Foundation [grant number 1553114 to S.D.S.]. The RMACC Summit and Alpine supercomputers used for the computational work are

561 also supported by the National Science Foundation [grant numbers ACI-1532235, ACI-1532236,
562 and ACI 2201538].
563
564 **Data Availability**
565 RNA-seq data from the backcross individuals have been uploaded to the SRA under Bioproject
566 PRJNA1092111. Data for the other six species have been uploaded to SRA Bioproject
567 PRJNA1102413. Code and additional data used in the analyses are available at
568 **https://osf.io/j5m8f/.**

References

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell,* **26**: 962-980.

Albert NW, Lewis DH, Zhang H, Schwinn KE, Jameson PE, Davies KM. 2011. Members of an R2R3-MYB transcription factor family in *Petunia* are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. *Plant Journal,* **65**: 771-84.

Auge GA, Penfield S, Donohue K. 2019. Pleiotropy in developmental regulation by flowering-pathway genes: is it an evolutionary constraint? *New Phytologist,* **224**: 55-70.

Ballester AR, Molthoff J, de Vos R, Hekkert B, Orzaez D, Fernández-Moreno JP, Tripodi P, Grandillo S, Martin C, Heldens J, Ykema M, Granell A, Bovy A. 2010. Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiology,* **152**: 71-84.

Berardi AE, Esfeld K, Jäggi L, Mandel T, Cannarozzi GM, Kuhlemeier C. 2021. Complex evolution of novel red floral color in *Petunia*. *Plant Cell,* **33**: 2273-2295.

Berardi AE, Hildreth SB, Helm RF, Winkel BS, Smith SD. 2016. Evolutionary correlations in flavonoid production across flowers and leaves in the Iochrominae (Solanaceae). *Phytochemistry,* **130**: 119-27.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**: 2114-20.

Bovy A, Schijlen E, Hall RD. 2007. Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. *Metabolomics,* **3**: 399-412.

Byers KJ, Vela JP, Peng F, Riffell JA, Bradshaw HD. 2014. Floral volatile alleles can contribute to pollinator-mediated reproductive isolation in monkeyflowers (*Mimulus*). *Plant Journal,* **80**: 1031-1042.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ,* **6**: e4958.

Casbon J. 2012. PyVCF - A Variant Call Format Parser for Python v. 0.3.0. 0.3.0 ed. https://github.com/jamescasbon/PyVCF.

Castillejo C, Waurich V, Wagner H, Ramos R, Oiza N, Muñoz P, Triviño JC, Caruana J, Liu Z, Cobo N, Hardigan MA, Knapp SJ, Vallarino JG, Osorio S, Martín-Pizarro C, Posé D, Toivainen T, Hytönen T, Oh Y, Barbey CR, Whitaker VM, Lee S, Olbricht K, Sánchez-Sevilla JF, Amaya I. 2020. Allelic variation of MYB10 is the major force controlling natural variation in skin and flesh color in Strawberry (*Fragaria* spp.) fruit. *The Plant Cell,* **32**: 3723-3749.

Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology,* **15**: 410.

Deanna R, Larter MD, Barboza GE, Smith SD. 2019. Repeated evolution of a morphological novelty: a phylogenetic analysis of the inflated fruiting calyx in the Physalideae tribe (Solanaceae). *American Journal of Botany,* **106**: 270-279.

Deanna R, González Ramírez I, Särkinen T, Knapp S, Sauquet H, Smith SD. Late Cretaceous origins for major nightshade lineages from total evidence timetree analysis. Invited submission for Special Issue on 'Milestones and trends: the role of the fossil record in reconstructing plant evolution' in *Annals of Botany*.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution,* **64**: 2044-2054.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* **29**: 15-21.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. *Trends in Plant Science,* **15**: 573-581.

Espley RV, Brendolise C, Chagné D, Kutty-Amma S, Green S, Volz R, Putterill J, Schouten HJ, Gardiner SE, Hellens RP, Allan AC. 2009. Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell,* **21**: 168-183.

Falginella L, Castellarin SD, Testolin R, Gambetta GA, Morgante M, Di Gaspero G. 2010. Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. *BMC Genomics,* **11**: 562.

Fattorini R, Ó'Maoiléidigh DS. 2022. *Cis*-regulatory variation expands the colour palette of the Brassicaceae. *Journal of Experimental Botany,* **73**: 6511-6515.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J,* **66**: 94-116.

Fernandez-Moreno JP, Tzfadia O, Forment J, Presa S, Rogachev I, Meir S, Orzaez D, Aharoni A, Granell A. 2016. Characterization of a new pink-fruited tomato mutant

results in the identification of a null allele of the SlMYB12 transcription factor. *Plant Physiology,* **171**: 1821-1836.

Gates DJ, Olson BJSC, Clemente TE, Smith SD. 2018. A novel R3 MYB transcriptional repressor associated with the loss of floral pigmentation in *Iochroma*. *New Phytologist,* **217**: 1346-1356.

Gates DJ, Strickler SR, Mueller LA, Olson BJ, Smith SD. 2016. Diversification of R2R3-MYB transcription factors in the tomato family Solanaceae. *Journal of Molecular Evolution,* **83**: 26-37.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* **29**: 644-52.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols,* **8**: 1494-1512.

Hileman LC. 2014. Trends in flower symmetry evolution revealed through phylogenetic and developmental genetic advances. *Philosophical Transactions of the Royal Society B,* **369**.

Holton TA, Brugliera F, Tanaka Y. 1993. Cloning and expression of flavonol synthase from *Petunia hybrida*. *Plant Journal,* **4**: 1003-1010.

Hopkins R, Rausher MD. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature,* **469**: 411-4.

Huang J, Xu W, Zhai J, Hu Y, Guo J, Zhang C, Zhao Y, Zhang L, Martine C, Ma H, Huang C-H. 2023. Nuclear phylogeny and insights into whole-genome duplications and reproductive development of Solanaceae plants. *Plant Communications,* **4**: 100595.

Jia D, Wu P, Shen F, Li W, Zheng X, Wang Y, Yuan Y, Zhang X, Han Z. 2021. Genetic variation in the promoter of an R2R3-MYB transcription factor determines fruit malate content in apple (*Malus domestica* Borkh.). *Plant Physiology,* **186**: 549-568.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution,* **30**: 772-80.

Kimura S, Koenig D, Kang J, Yoong FY, Sinha N. 2008. Natural variation in leaf morphology results from mutation of a novel KNOX gene. *Current Biology,* **18**: 672-677.

Koski MH, Ashman T-L. 2014. Dissecting pollinator responses to a ubiquitous ultraviolet floral pattern in the wild. *Functional Ecology,* **28**: 868-877.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* **9**: 559.

Larter M, Dunbar-Wallis A, Berardi AE, Smith SD. 2018. Convergent evolution at the pathway Level: predictable regulatory changes during flower color transitions. *Molecular Biology and Evolution,* **35**: 2159-2169.

Larter M, Dunbar-Wallis A, Berardi AE, Smith SD. 2019. Developmental control of convergent floral pigmentation across evolutionary timescales. *Developmental Dynamics,* **248**: 1091-1100.

Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. *Evolution,* **62**: 2131-2154.

Lüthi MN, Berardi AE, Mandel T, Freitas LB, Kuhlemeier C. 2022. Single gene mutation in a plant MYB transcription factor causes a major shift in pollinator preference. *Current Biology,* **32**: 5295-5308.e5.

Martins TR, Jiang P, Rausher MD. 2017. How petals change their spots: cis-regulatory re-wiring in Clarkia (Onagraceae). *New Phytologist,* **216**: 510-518.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology,* **14**: e1005944.

Michalak P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics,* **91**: 243-248.

Millar AA, Gubler F. 2005. The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell,* **17**: 705-721.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution,* **37**: 1530-1534.

Mol J, Grotewold E, Koes R. 1998. How genes paint flowers and seeds. *Trends in Plant Science,* **3**: 212-217.

Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, Timoneda A, Yim WC, Cushman JC, Yang Y. 2021. Disentangling sources of gene

tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology,* **70**: 219-235.

Mu Q, Wei J, Longest HK, Liu H, Char SN, Hinrichsen JT, Tibbs-Cortes LE, Schoenbaum GR, Yang B, Li X, Yu J. 2024. A MYB transcription factor underlying plant height in sorghum qHT7.1 and maize Brachytic 1 loci. *Plant Journal,* **120**: 2172-2192.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution,* **32**: 268-274.

Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant Physiol,* **171**: 2294-2316.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods,* **14**: 417-419.

Pollak PE, Vogt T, Mo Y, Taylor LP. 1993. Chalcone synthase and flavonol accumulation in stigmas and anthers of *Petunia hybrida*. *Plant Physiology,* **102**: 925-932.

Powell AF, Zhang J, Hauser D, Vilela JA, Hu A, Gates DJ, Mueller LA, Li FW, Strickler SR, Smith SD. 2022. Genome sequence for the blue-flowered Andean shrub *Iochroma cyaneum* reveals extensive discordance across the berry clade of Solanaceae. *Plant Genome,* **15**: e20223.

Preston JC, Martinez CC, Hileman LC. 2011. Gradual disintegration of the floral symmetry gene network is implicated in the evolution of a wind-pollination syndrome. *Proceedings of the National Academy of Sciences USA,* **108**: 2343-8.

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature,* **440**: 1050-1053.

Quattrocchio F, Verweij W, Kroon A, Spelt C, Mol J, Koes R. 2006. PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell,* **18**: 1274-1291.

Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R. 1999. Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *Plant Cell,* **11**: 1433-44.

Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci,* **10**: 63-70.

Rausher MD. 2006. The evolution of flavonoids and their genes. In: Grotewold E, ed. *The science of flavonoids*. New York, NY: Springer New York.

Rosa-Martínez E, Bovy A, Plazas M, Tikunov Y, Prohens J, Pereira-Dias L. 2023. Genetics and breeding of phenolic content in tomato, eggplant and pepper fruits. *Frontiers in Plant Science,* **14**: 1135237.

Ryan KG, Swinny EE, Winefield C, Markham KR. 2001. Flavonoids and UV photoprotection in *Arabidopsis* mutants. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences,* **56**: 745-754.

Sagawa JM, Stanley LE, LaFountain AM, Frank HA, Liu C, Yuan YW. 2016. An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers. *New Phytologist,* **209**: 1049-1057.

Schilbert HM, Glover BJ. 2022. Analysis of flavonol regulator evolution in the Brassicaceae reveals MYB12, MYB111 and MYB21 duplications and MYB11 and MYB24 gene loss. *BMC Genomics,* **23**: 604.

Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell,* **18**: 831-851.

Schwinn KE, Boase MR, Bradley JM, Lewis DH, Deroles SC, Martin CR, Davies KM. 2014. MYB and bHLH transcription factor transgenes increase anthocyanin pigmentation in petunia and lisianthus plants, and the petunia phenotypes are strongly enhanced under field conditions. *Frontiers in Plant Science,* **5**: 603.

Sheehan H, Moser M, Klahre U, Esfeld K, Dell'Olivo A, Mandel T, Metzger S, Vandenbussche M, Freitas L, Kuhlemeier C. 2016. MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. *Nature Genetics,* **48**: 159-166.

Shiu SH, Shih MC, Li WH. 2005. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiology,* **139**: 18-26.

Singh P, Arif Y, Bajguz A, Hayat S. 2021. The role of quercetin in plants. *Plant Physiology and Biochemistry,* **166**: 10-19.

Smith SD, Baum DA. 2006. Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). *American Journal of Botany,* **93**: 1140-53.

Smith SD, Hall SJ, Izquierdo PR, Baum DA. 2008. Comparative pollination biology of sympatric and allopatric Andean *Iochroma* (Solanaceae). *Annals of the Missouri Botanical Garden,* **95**: 600-617.

Smith SD, Rausher MD. 2011. Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution,* **28**: 2799-2810.

Smith SD, Wang S, Rausher MD. 2013. Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Molecular Biology and Evolution,* **30**: 602-612.

Sobel JM, Streisfeld MA. 2013. Flower color as a model system for studies of plant evo-devo. *Frontiers in Plant Science,* **4**: 321.

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research,* **4**: 1521.

Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience,* **4**: 48.

Song Z, Luo Y, Wang W, Fan N, Wang D, Yang C, Jia H. 2019. NtMYB12 positively regulates flavonol biosynthesis and enhances tolerance to low Pi stress in *Nicotiana tabacum*. *Frontiers in Plant Science,* **10**: 1683.

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J,* **50**: 660-77.

Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology,* **4**: 447-456.

Streisfeld MA, Young WN, Sobel JM. 2013. Divergent selection drives genetic differentiation in an R2R3-MYB transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLoS Genet,* **9**: e1003385.

Todesco M, Bercovich N, Kim A, Imerovski I, Owens GL, Dorado Ruiz Ó, Holalu SV, Madilao LL, Jahani M, Légaré JS, Blackman BK, Rieseberg LH. 2022. Genetic basis and dual adaptive role of floral pigmentation in sunflowers. *Elife,* **11**.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. 2013. The evolution of phenylpropanoid metabolism in the green lineage. *Critical Reviews in Biochemistry and Molecular Biology,* **48**: 123-152.

Wessinger CA, Rausher MD. 2012. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany,* **63**: 5741-5749.

Wessinger CA, Rausher MD. 2015. Ecological transition predictably associated with gene degeneration. *Molecular Biology and Evolution,* **32**: 347-354.

Wheeler LC, Dunbar-Wallis A, Schutz K, Smith SD. 2023. Evolutionary walks through flower colour space driven by gene expression in *Petunia* and allies (Petunieae). *Proceedings of the Royal Society B,* **290**: 20230275.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology,* **126**: 485-493.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet,* **8**: 206-16.

Wu Y, Popovsky-Sarid S, Tikunov Y, Borovsky Y, Baruch K, Visser RGF, Paran I, Bovy A. 2023. CaMYB12-like underlies a major QTL for flavonoid content in pepper (*Capsicum annuum*) fruit. *New Phytologist,* **237**: 2255-2267.

Wu Y, Wen J, Xia Y, Zhang L, Du H. 2022. Evolution and functional diversification of R2R3-MYB transcription factors in plants. *Horticulture Research,* **9**: uhac058.

Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution,* **31**: 3081-3092.

Yarahmadov T, Robinson S, Hanemian M, Pulver V, Kuhlemeier C. 2020. Identification of transcription factors controlling floral morphology in wild *Petunia* species with contrasting pollination syndromes. *Plant Journal,* **104**: 289-301.

Yuan YW, Rebocho AB, Sagawa JM, Stanley LE, Bradshaw HD. 2016. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proceedings of the National Academy of Sciences USA,* **113**: 2448-2453.

Zheng Y, Chen Y, Liu Z, Wu H, Jiao F, Xin H, Zhang L, Yang L. 2021. Important roles of key genes and transcription factors in flower color differences of *Nicotiana alata*. *Genes (Basel),* **12**: 1976.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics,* **44**: 821-824.

Table 1. Phenotypes and genotypes of sampled individuals from backcross population. The DEL, CYAN and PEL columns show the proportion of anthocyanins derived from blue delphinidin, purple cyanidin and red pelargonidin pigments, respectively (data from Smith and Rausher 2011). The expression of *F3'h* was quantified with qPCR in Smith and Rausher (2011); individuals with 'low' expression have 10-fold lower expression than those with 'high'. Individuals with high *F3'h* expression and/or primarily cyanidin production are predicted to be heterozygous at the *T*-locus with one 'blue' and one 'red' allele (Tt). The samples are split between *Tt* and *tt* individuals at the *T*-locus, and there are three replicates for each combination of genotypes at the other involved loci (*F3'5'h* and *Dfr*). Note that the red parental species is missing the functional copy of *F3'5'h,* so the red allele is indicated with a -.

Indiv	DEL	CYAN	PEL	*F3'h* expression	Inferred T-locus genotype	*F3'5'h*	*Dfr*
GCG22	0.2	0.68	0.1	high	***Tt***	***F-***	*dd*
GCG55	0.3	0.63	0.1	high	***Tt***	***F-***	*dd*
GCG11	0.24	0.61	0.15	high	***Tt***	***F-***	*dd*
GCG98	0.2	0.7	0.1	high	***Tt***	–	***Dd***
GCG84	0.03	0.67	0.30	high	***Tt***	–	***Dd***
GCG25	0.2	0.62	0.2	high	***Tt***	–	***Dd***
GCG49	0.12	0.53	0.35	high	***Tt***	***F-***	***Dd***
GCG40	0.11	0.67	0.22	high	***Tt***	***F-***	***Dd***
GCG94	0.3	0.57	0.2	n/a	***Tt***	***F-***	***Dd***
GCG60	0.01	0.70	0.29	high	***Tt***	–	*dd*
GCG18	0.01	0.84	0.14	high	***Tt***	–	*dd*
GCG76	0.05	0.81	0.14	n/a	***Tt***	–	*dd*
GCG2	0.24	0.07	0.69	low	*tt*	***F-***	***Dd***
GCG61	0.2	0.14	0.7	low	*tt*	***F-***	***Dd***
GCG23	0.21	0.11	0.67	low	*tt*	***F-***	***Dd***
GCG24	0.2	0.14	0.7	low	*tt*	***F-***	*dd*
GCG73	0.2	0.12	0.7	low	*tt*	***F-***	*dd*
GCG7	0.17	0.14	0.69	low	*tt*	***F-***	*dd*
GCG4	0.05	0.05	0.90	low	*tt*	–	***Dd***
GCG85	0.02	0.04	0.94	n/a	*tt*	–	***Dd***
GCG6	0.08	0.11	0.81	low	*tt*	–	***Dd***
GCG9	0.05	0.04	0.91	low	*tt*	–	*dd*
GCG104	0.1	0.08	0.9	low	*tt*	–	*dd*
GCG43	0.07	0.06	0.87	low	*tt*	–	*dd*

Figures and figure legends

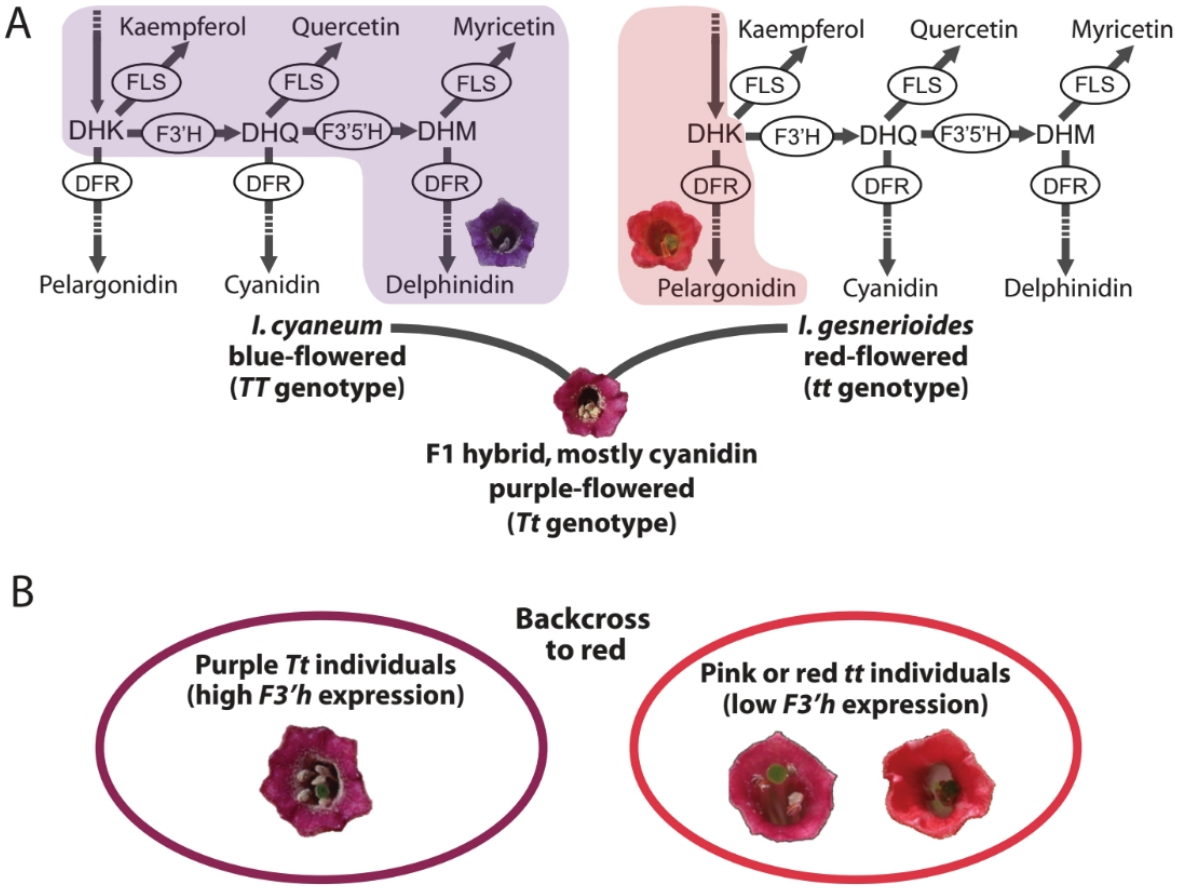

Fig. 1. Flavonoid pigment production in parental lines and experimental design for identifying the T-locus. (A) Segregating backcross populations were created from parental lines of the blue-flowered *Iochroma cyaneum* and the red-flowered *I. gesnerioides*. The former makes delphinidin-derived anthocyanins and all three of flavonols (kaempferol, quercetin and myricetin) while the latter makes only pelargonidin-derived anthocyanins and kaempferol. The active branches of the pathway are shaded in each case. The enzymes shown (in ellipses) are flavonoid 3'hydroxylase (F'3H), flavonoid-3'5'-hydroxylase (F3'5'H), dihydroflavonol reductase (DFR) and flavonol synthase (FLS). Flavonoid intermediates are dihydrokaempferol (DHK), dihydroquercetin (DHQ) and dihydromyricetin (DHM). Additional steps upstream of DHK (e.g. involving chalcone synthase, chalcone isomerase and flavanone hydroxylase) and downstream of DFR (e.g., involving anthocyanidin synthase, glucosyltransferase) are not shown but indicated with the dashed portion of the arrows. Note that F3'5'H has 3' activity and can act on DHK in some taxa, but in *Iochroma*, it is specialized for DHQ (Smith and Rausher, 2011). The F1 hybrid produces mainly cyanidin-derived anthocyanins and is presumed to be heterozygous at the *T*-locus, which controls *F3'h* expression. (B) Phenotypes and pools for RNASeq experiment. We divided the backcross population (F1 crossed to the red parent) into a high *F3'h* expression purple-flowered cyanidin-producing pool (presumably *Tt*) and a low *F3'h* expression mostly or entire pelargonidin-producing pink to red-flowered pool (presumably *tt*). See Table 1 for more information on sequenced individuals.

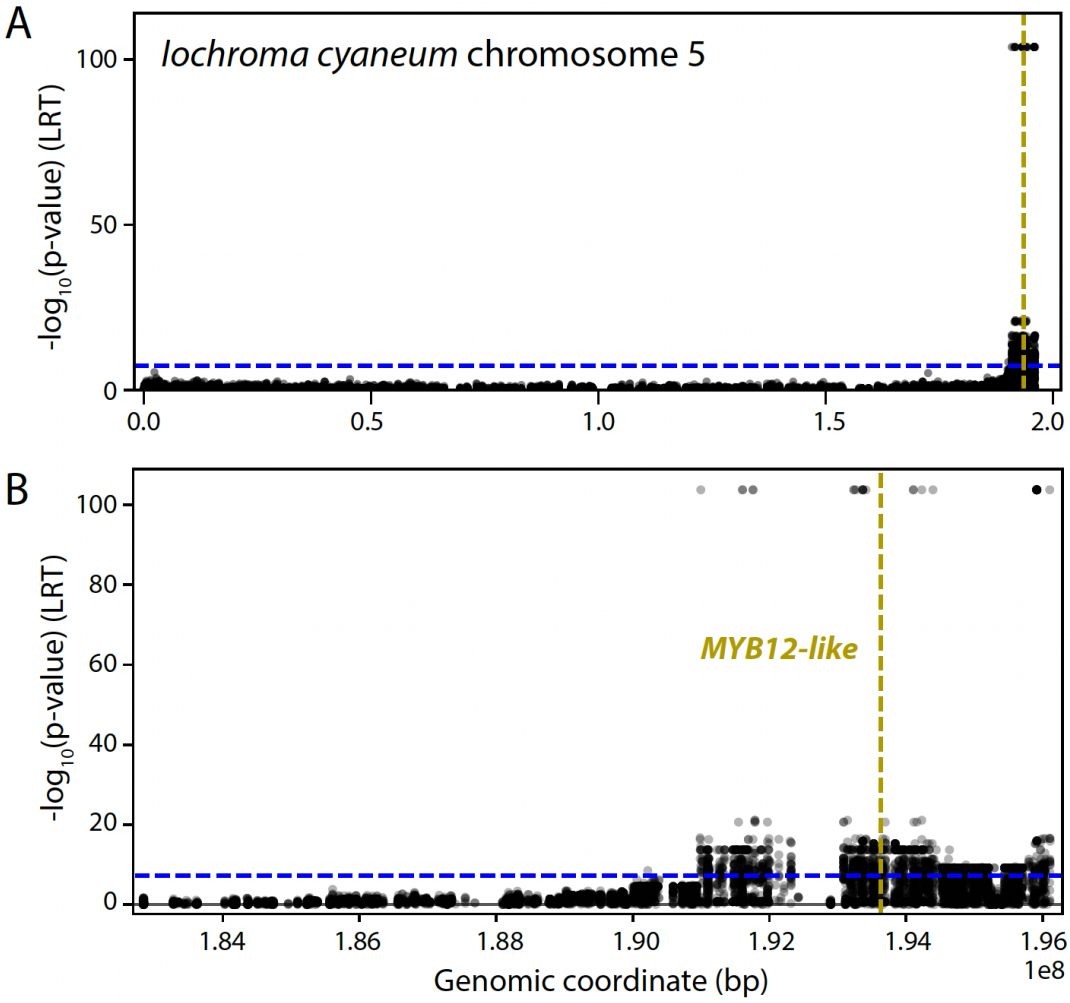

Fig. 2. Manhattan plot showing significant SNP associations on *Iochroma cyaneum* chromosome 5. The blue-dashed line marks the cutoff for genome-wide significance ($P<5x10^{-8}$). The gold vertical line marks the location of the *MYB12-like* gene, which predicts *F3'h* expression. (A) The position of *MYB12-like* near the 3' prime end of the chromosome. (B) Close-up of the region containing *MYB12-like*, showing the concentration of associated SNPS in the last 500kb of the chromosome.

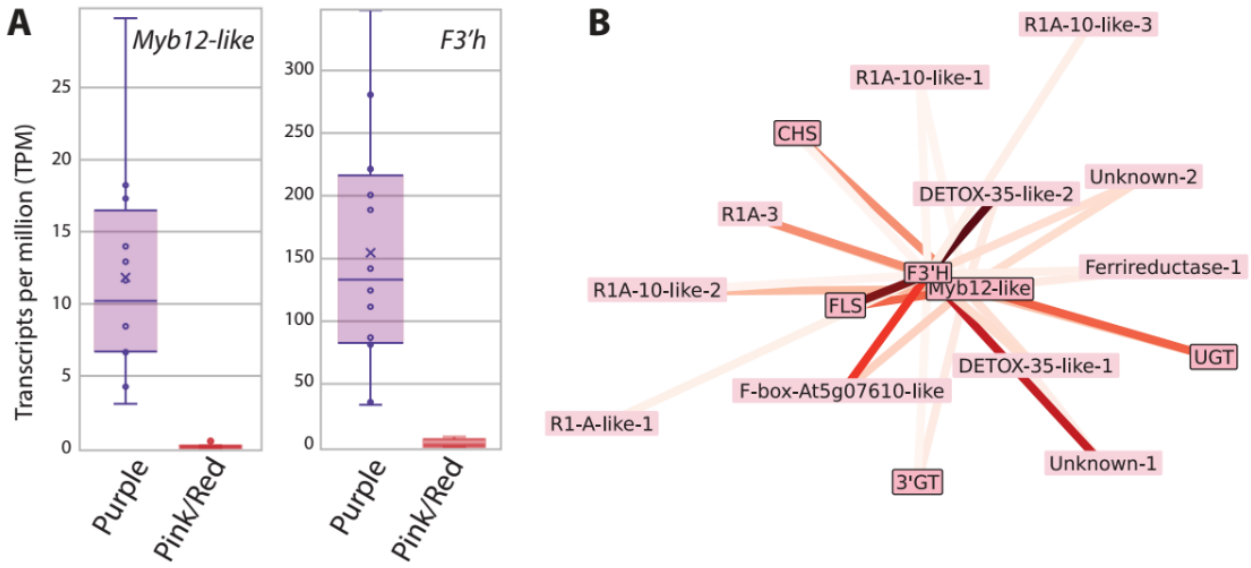

Fig. 3. Co-expression of *F3'h* and *IcMYB12-like*. (A) Expression levels for each gene in the two phenotypic pools. Box plots mark the first and third quantiles, with a bisecting line to indicate the median. The mean is denoted with an x. The Pearson correlation coefficient for the expression of these two genes is 0.91 (P<0.0001). (B) Submodule from WGCNA analysis containing all edges including *F3'h* and *IcMYB12-like* (see Supplementary Table S4). The lines representing each edge are colored by the connectivity value (TOM) from WGCNA (Fig. S6); more closely clustered genes are more tightly co-expressed (i.e., spring layout). Enzymes related to the flavonoid pathway (CHS, FLS, 3'GT, UGT) are outlined along with *IcMYB12-like*.

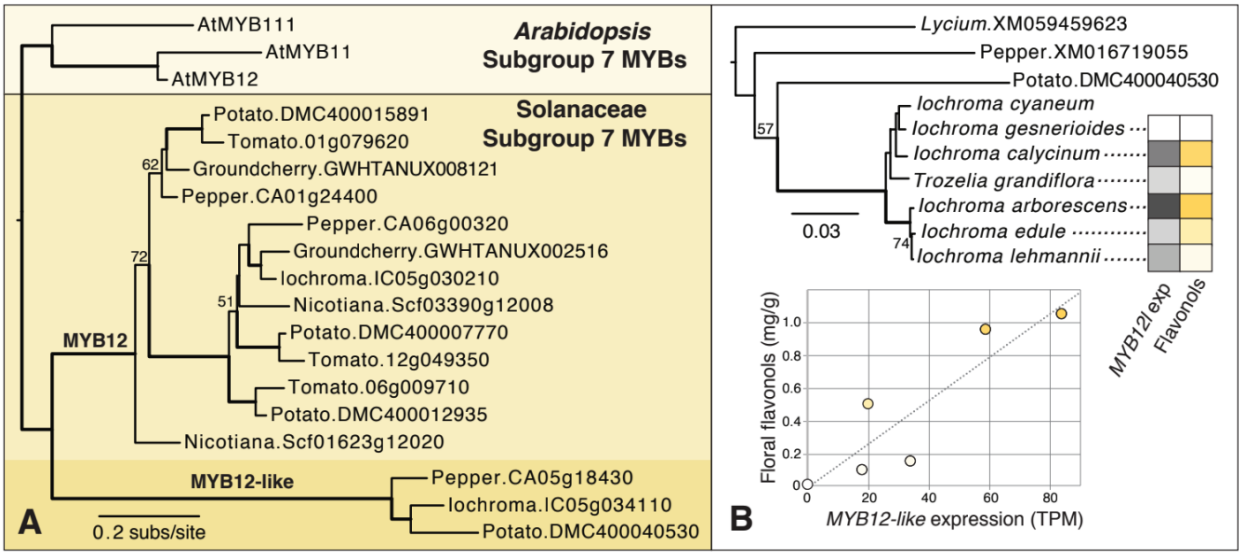

Fig. 4. Phylogenetic position of MYB12-like proteins in relation to other subgroup 7 R2R3 MYBs. (A) Maximum likelihood phylogeny from protein sequences. Most Solanaceae subgroup 7 MYBs fall into large clade typically annotated as “MYB12”. MYB12-like sequences fall into a deeply diverged clade that appears to be sister to the MYB12 sequences. Bolded branches have >95% bootstrap support; values between 50 and 95% bootstrap support are shown. (B) Maximum likelihood phylogeny for MYB12-like sequences based on a complete CDS alignment. Bootstrap supports are shown as in (A). Tip values for *MYB12-like* expression (TPM) and floral flavonol content (in mg/g from Larter et al. 2019) for six species are colored by magnitude (see Supplementary Table S5 for raw data). These data are graphed in the inset figure with the dashed line showing the linear trend. *Iochroma cyaneum* is not included as data from previous transcriptomic analyses (Gates et al. 2018) are not directly comparable with the de novo transcriptomes from the present study. Full names and sources for all sequences used in this analysis are given in Supplementary Table S6. Branch lengths in both trees are in substitutions per site.