A novel R3 MYB transcription repressor associated with the loss of floral pigmentation in Iochroma

Daniel J. Gates[1] , Bradley J. S. C. Olson[2], Tom E. Clemente[3] and Stacey D. Smith[4]

[1]School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588, USA;
[2]Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA; [3]Department of Agronomy and Horticulture and Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska 68588, USA;
[4]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder Colorado, 80305, USA

Authorfor correspondence:
Stacey D. Smith
Tel: +1 3034921374
Email: stacey.d.smith@colorado.edu
Twitter: @iochromaland

Received: July 6, 2017
Accepted: August 31, 2017

Key words: anthocyanins, EAR motif, flower color, gene duplication, transcription factor

Total word count: 5290
Introduction: 587
Materials and Methods: 1512
Results: 1978
Discussion: 1213
Acknowledgements: 59
Number of Figures: 6
Number of Tables: 0
Supporting Information: 3 Tables, 2 Figures

Summary

- Losses of floral pigmentation represent one of the most common evolutionary transitions in flower color, yet the genetic basis for these changes has been elucidated in only a handful of cases.
- Here we used crossing studies, bulk-segregant RNA sequencing, phylogenetic analyses and functional tests to identify the gene(s) responsible for the transition to white flowers in **Iochroma loxense**.
- Crosses between **I. loxense** and its blue-flowered sister species, **I. cyaneum**, suggested that a single locus controls the flower color difference and that the white allele causes a nearly complete loss of pigmentation. Examining sequence variation across phenotypic pools from the crosses, we found that

alleles at a novel R3 MYB transcription factor were tightly associated with flower color variation. This gene, which we term **MYBL1**, falls into a class of MYB transcriptional repressors, and accordingly, higher expression of this gene is associated with downregulation of multiple anthocyanin pigment pathway genes. We confirmed the repressive function of **MYBL1** through stable transformation of **Nicotiana**.

- The mechanism underlying the evolution of white flowers in **I. loxense** differs from that uncovered in previous studies, pointing to multiple mechanisms for achieving fixed transitions in flower color intensity.

Introduction

From a genetic and biochemical perspective, the biosynthesis of anthocyanin pigment represents one of the best studied metabolic pathways in plants. The red, purple, and blue anthocyanins are the products of a series of enzymatic steps that are highly conserved across plants, and they constitute the most common pigments responsible for flower and fruit coloration (Winkel-Shirley, 2001; Grotewold, 2006). The regulation of anthocyanin production varies in different taxa, but members of the WD-repeat (WDR), basic helix-loop-helix (bHLH) and MYB families are commonly important regulators of this pathway (Feller **et al.**, 2011; Davies **et al.**, 2012). Often, these transcription factors form a complex (the MYB-bHLH-WDR or MBW complex) that can coordinately activate or repress multiple steps in the pathway to modulate pigment production (Albert **et al.**, 2014).

This deep understanding of anthocyanin pathway function and regulation has made it possible to identify the genetic basis of pigment variation in many non-model plants. These studies have begun to reveal predictable patterns in pigment evolution (Streisfeld & Rausher, 2011; Sobel & Streisfeld, 2013). For example, evolutionary transitions to white flowers via losses of floral anthocyanin production have consistently involved loss-of-function mutations in R2R3 MYB transcriptional activators (Quattrocchio **et al.**, 1999; Schwinn **et al.**, 2006; Hoballah **et al.**, 2007). Given that a wide range of pathway mutations can give rise to white flowers (de Vlaming **et al.**, 1984; van Houwelingen **et al.**, 1998), this pattern has been attributed to preferential fixation of these R2R3 MYB mutations (Streisfeld & Rausher, 2011). R2R3 MYB transcription factors are among the largest gene families in plants, and each member is highly specialized in terms of spatial expression, timing of expression, and target genes (Kranz **et al.**, 1998; Ramsay & Glover, 2005). Thus, mutations in R2R3 MYBs have the

potential to change anthocyanin gene expression only in the flower with minimal pleiotropic consequences on the production of anthocyanins or related flavonoids in other tissues (Wessinger & Rausher, 2012). Nonetheless, this apparent preferential fixation is based on studies of floral pigment loss in two genera (**Antirrhinum** and **Petunia**) and remains to be validated more broadly.

Here we examine the genetic basis for the transition to white flowers in the Andean genus **Iochroma** (Solanaceae). **Iochroma** is a relatively small genus of around 25 species, but presents a wide range of flower colors, including red, blue, green, white, yellow and orange (Smith & Baum, 2006). The ancestral state in the genus is blue flowers, which derive their color from delphinidin-based anthocyanins (Smith & Baum, 2007; Berardi **et al.**, 2016). From this ancestral state, the clade has experienced multiple transitions, leading to a variety of derived color phenotypes, including at least three to white or yellow flowers due to losses of floral anthocyanins (Smith & Goldberg, 2015). The present study focuses on the most recent of these transitions, along the branch leading to the white-flowered **I. loxense**. Its sister species, **I. cyaneum**, is a horticulturally important plant in which the anthocyanin pathway and anthocyanin production have been previously characterized (Smith & Rausher, 2011; Berardi **et al.**, 2016). Taking advantage of the crossability of these two taxa, segregating backcross populations were created to identify genes associated with flower color variation. Given the large number of potential candidate genes, including the many R2R3 MYBs, we adopted a bulk-segregant RNA-Seq approach to identify potentially causative loci and determine their effects on pathway gene expression. The results implicate a novel class of R3 MYB transcriptional repressors, which appear to influence flower color via the same developmental changes as the previously identified R2R3 loss-of-function mutations.

Materials and Methods

Crossing and phenotyping

We created a cross between the blue-flowered **I. cyaneum** andthe white-flowered **I. loxense** using two cultivated accessions. The parental **I. cyaneum** individual was derived from an accession cultivated at the Missouri Botanical Garden by W. G. D'Arcy, and this accession was used to generate a low-coverage genome in previous work (Gates **et al.**, 2016). The parental **I. loxense** individual was derived from material from the University of Loja Botanical Gardens near Loja, Ecuador. Voucher specimens for each accession (Smith 265 and 235, respectively) are deposited at the University of Wisconsin-Madison Herbarium and the Missouri Botanical Garden. Previous biochemical studies confirm that **I. cyaneum** produces over 97% blue delphinidin-based anthocyanin pigments, while **I. loxense** produces no detectable floral anthocyanins (Berardi **et al.**, 2016). A single F1 individual was backcrossed to each parental accession to create segregating backcross populations. Backcross individuals were grown in greenhouse conditions with natural light. Flowers from each backcross individual were photographed, and individuals were visually grouped into three phenotypic classes (blue-, white- and F_1-phenotype). Floral anthocyanins were quantified by eluting pigments from fresh corolla tissue in 500 µl of methanolic HCl (1%) and measuring absorbance at 530nm. Reflectance spectra were collected from the corolla tube and lobe of each individual using a JAZ spectrometer as described in Ng and Smith (2016). Spectral variation was examined by principal component analysis of the combined corolla tube and lobe data in the visible spectrum (400-700nm).

Association mapping

Bulk transcriptome sequencing of each parent and phenotypic pool was used to search for allelic variants associated with flower color. Floral bud RNA was extracted following Coburn et al. (2015) from the blue and white parents (two biological replicates each); 20 blue- and 15 F_1-phenotypeindividuals from the backcross to the blue parent; and 8 white- and 9 F_1-phenotype individuals from the backcross to the white parent. Two lanes of Illumina 100-base-pair paired-end RNAseq of Tru-Seq libraries (Illumina, Inc., San Diego, CA) were generated for these six pools of RNA. We created a transcriptome index from the **I. cyaneum** transcriptome (Gates **et al.**, 2016) with Bowtie 2.02 (Langmead **et al.**, 2009) with default settings and mapped reads from each pool using TopHat 2.1.0 (Kim **et al.**, 2013) allowing 5 mismatches per read (-N flag) and a final read edit distance of 8. This TopHat mapping for each pool generated a BAM file to be used in both association and expression analyses. In addition to each TopHat BAM mapping file, downstream expression and variant calling analyses also require a gtf mapping file. Without a prespecified gtf from an annotated genome assembly, TopHat makes run-specific gtf mapping files along with each BAM mapping but can exclude low expression genes of interest (e.g. anthocyanin pathway genes in white-flowered pools) in these gtf files. To avoid potentially missing genes in the gtf file, we used BWA-MEM (Li & Durbin, 2009) to construct a pan-gtf file by mapping reads from all libraries to the **I. cyaneum** transcriptome. This mapping was passed to TopHat for gtf file construction and the resulting gtf file contained all gene models with reads that were mapped in any pool. We then used SAMtools (Li **et al.**, 2009) to generate variant calls in each of the mappings. Before association mapping, we removed low quality variants with less than 10-fold coverage and with a quality score of less than 50.

Two approaches were used to test for associations between genes and color phenotypes using the mapped reads. First, we conducted a transcriptome-wide search for genes with patterns and frequencies of SNP variation consistent with the phenotypic pools. We focused on genes that differed between the parents by 3 or more SNPs with 200x coverage to increase confidence in the assignment of alleles to the **I. cyaneum** or **I. loxense** parents. We filtered this initial list of genes, saving those which showed no mapped SNPs in the blue backcross pool (i.e. all reads match the blue **I. cyaneum** parent). This follows from the expectation that blue-flowered individuals should be homozygous for the blue parental allele at the causative locus. We further reduced the pool by applying a second filter, keeping only genes with 35 to 65% blue variants in the F_1 phenotypic pools and less than 35% blue variants in the white phenotypic pool. With precise phenotyping, we would expect 0% blue variants in the white pool; however, distinguishing the two phenotypes was difficult in the backcross to the white parent. We expected the F1 phenotypic pool to have intermediate allele frequency, but not precisely 50% blue variants because of the difficulty in phenotyping as well as any cis-regulatory effects.

Next, we compared loci recovered from this transcriptome-wide pipeline to SNP variation at candidate loci in the anthocyanin pathway. We included seven structural genes (**Chs**, **Chi**, **F3h**, **F3'h**, **F3'5'h**, **Dfr**, **Ans**) and six known transcription factors (the R2R3 MYB activator **AN2** and repressor **MYB27**, the bHLH genes **jaf13** and **AN1**, the WDR **AN11**, and the R3 MYB repressor **MYBx**) (Koes **et al.**, 2005; Albert **et al.**, 2014). We refer to the names of the loci in **Petunia hybrida** for six transcription factors (**AN1**, **AN2**, etc.) for ease of comparison with existing literature. The **Iochroma** orthologs for these pathway genes have been characterized in previous studies (Smith & Rausher, 2011; Coburn **et al.**, 2015; Gates **et al.**, 2016). For each locus, we calculated the proportion of reads matching the blue **I. cyaneum** parent allele in each phenotypic pool.

Phylogenetic analyses of MYBL1

A combination of bioinformatic and phylogenetic analyses were used to elucidate the evolutionary history of **MYBL1**, an R3 MYB gene identified by our pipeline as having a strong association with flower color. Potential orthologs in related taxa were retrieved through a combination of BLAST searches and gene prediction using Augustus (Stanke & Morgenstern, 2005). Genome assemblies for tomato (**Solanum lycopersicum**), potato (**S. tubersosum**), **S. pennellii,** pepper (**Capsicum annuum**), tobacco (**Nicotiana tabacum**) and petunia (**Petunia hybrida**), and coffee (**Coffea canephora**) were accessed through the Sol Genomics Network (solgenomics.net) and the Coffee Genome Hub (coffee-genome.org). Protein sequences for MYBL1 were aligned to similar Solanaceae MYB sequences along with representatives of the major MYB subgroups in **Arabidopsis** (Stracke **et al.**, 2001; Dubos **et al.**, 2010) for phylogenetic analysis. The sequences were aligned using Clustal in Geneious 6.05 (Biomatters Inc., Auckland, NZ) with default settings. The phylogeny was estimated in MrBayes 3.2.5 (Ronquist & Huelsenbeck, 2003) with 2 independent runs, each of 1 million generations, and a mixed prior on models. Convergence was judged by PSRF values approaching 1.0, estimated sample sizes (ESS) exceeding 300, and similarity of the consensus trees from the two runs.

Expression analyses

Expression levels of anthocyanin transcription factors and regulatory genes in each of our six RNA pools (the two parents and the two phenotypes in each backcross population) were measured by mapping reads to the **I. cyaneum** transcriptome as described above. After mapping we calculated FPKM and tested for differential expression using Cuffdiff 1.1.2 (Trapnell **et al.**, 2010) with our custom .gtf file. Cuffdiff uses an optimization routine to obtain a maximum a priori estimate of the FPKM for a given fragment (typically a gene). From this sampling routine, the program also calculates a 95% confidence interval based upon the variance-covariance relationship of high likelihood samples that are also weighted by an information index (as a means of ensuring lower coverage fragments will have appropriate levels of uncertainty). We used the FPKM estimates and accompanying confidence intervals to compare expression levels across pools for each gene by using a two-sample t-test where thestandard error for the expression of each gene at each pool is calculated as the confidence interval divided by 1.96. We also conducted quantitative real-time PCR(qPCR) to confirm patterns of expression across the phenotypic pools. Three individuals were randomly selected from the four backcross pools. We completed two replicate qPCR reactions for each individual and three for each parent for seven structural genes and four transcription factors (see Supporting Information Table S1 for primers). Two of the six transcription factors included in the association analysis (**MYB27** and **AN1**) were excluded from the qPCR experiment because of their low expression across all pools in the RNASeq experiment. Protocols for cDNA synthesis,qPCR reaction conditions, and data analysis follow Coburn et al. (2015).

Functional testing

The function of **IlMYBL1** was tested through stable transformation of **N. tabacum**. First, the complete coding sequence of **MYBL1** from **I. loxense** was synthesized (GenScript Corp., Piscataway, NJ) and cloned into the vector pRTL2 vector (Carrington & Freed, 1990), resulting in constitutive expression regulated by the enhanced cauliflower mosaic virus 35S (e35S) promoter with the tobacco etch virus (TEV) translational enhancer element. The resultant expression cassette was subsequently subcloned into the binary vector pPXP212 (Hajdukiewicz **et**

al., 1994). Tobacco leaf disks were transformed with this vector following the protocol outlined in Clemente (2006), and 20 of the resulting transformants were grown seed in the greenhouse. Floral pigmentation was quantified from both transgenic and wild-type plants by eluting anthocyanins from petal disks with 500 µl of methanolic HCl (1%) and measuring absorbance at 530nm.

Results

Patterns of phenotypic variation in crosses suggest a single underlying locus

Crosses between the white-flowered **Iochroma loxense** and the blue-flowered **I. cyaneum** resulted in a pale (nearly white) phenotype. The F_1 has a white floral tube with slight anthocyanin pigmentation near the mouth of the corolla (Fig. 1a). Such patterning is not apparent in the parents, which are uniformly colored throughout the tube and the mouth. The highly reduced level of floral pigmentation in the F_1 suggests nearly complete dominance of the allele (or alleles) from the white-flowered **I. loxense.**

Backcrosses of the F_1 to **I. loxense** and **I. cyaneum** recovered the three parental phenotypes, consistent with a single major locus determining the flower color difference**.** Both blue and F_1 phenotypes segregated in the backcross to the blue **I. cyaneum** parent, and F_1 and white phenotypes in the backcross to the white **I. loxense** parent (Fig. 1a). As expected for a single gene controlling the phenotype, the numbers of individuals with each phenotype did not significantly differ from 1:1 (backcross to blue, 35 blue: 38 F_1 phenotype; backcross to white, 8 white: 11 F_1 phenotype, P=0.8, Fisher's exact test). Fewer individuals were scored in the backcross to white because of low viability in that cross (possibly due to incipient incompatibility between the sister species).

Quantitative comparisons of pigment concentration showed a clear split between the blue phenotype and the other two phenotypes. The blue-flowered parent and the blue-flowered individuals from the backcross to the blue parent produced roughly 5 times the amount of anthocyanin pigment as the white or F_1 -phenotype individuals, with no overlap in values (Fig. 1b). Although the white-flowered backcross individuals had lower average pigment content than the F_1 -phenotype individuals, there was significant overlap among the pools (Fig. 1b).

A similar pattern was observed in a principal component analysis of floral color spectra, which sharply divided the blue-flowered individuals and the rest. The first principal component axis captured 84% of the variation and split the individuals into two clusters, **I. cyaneum** and the blue-flowered individuals from the backcross to **I. cyaneum** versus the others (Supporting Information Fig. S1a). This axis corresponds to brightness (the total amount of light reflected), as each of the wavelengths of light has roughly equal loading (Supporting Information Fig. S1b). As expected, blue flowers have low brightness, while the white and F_1 -phenotype flowers have high brightness (Supporting Information Fig. 1a). The other minor PC axes incorporate relative reflectance across different wavelengths (floral hue), and here, indicate different reflectance in green/yellow wavelengths vs. red/blue (Supporting Information Fig. S1b).

Bulk-segregant mapping points to an R3 MYB associated with color

We used a bulk-segregant RNA-Seq approach to identify genes associated with the phenotypic classes from the crosses. Although the white and F_1 phenotypes overlap in pigment concentration and reflectance spectra, we sequenced each of the pools separately in order to detect expression differences that could account for their distinct color patterns (Fig. 1a). Our first filtering of the transcriptomic pools for associated genes (requiring

three confidently called SNPs between the parents and no variants called between **I. cyaneum** and the blue-flowered pool) resulted in 35 candidates (Supporting Information Table S2). Among these candidates, only one belonged to a class of genes (MYB transcription factors) known to regulate anthocyanin production. Although the gene appeared to be a single repeat R3 MYB, the closest BLAST hit in tomato was an R2R3 MYB with uncharacterized function (Supporting Information Table S2). We next filtered the list of 35 candidates by SNP frequency in the remaining pools, retaining genes with 35 to 65% reads matching **I. cyaneum** in the F_1 phenotype pools and less than 35% in the white pool. The apparent R3 MYB sequence was the only gene to pass this second filter, and given its phylogenetic proximity to **A. thaliana MYBL2** (see below), this locus was designated **MYBlike-1** or **MYBL1**. The **MYBL1** sequences for **I. loxense** and **I. cyaneum** and have been uploaded to Genbank as KY658468 and KY658469.

We compared these results to the known structural and regulatory genes of anthocyanin pathway, none of which emerged from the transcriptome-wide search. All of the structural genes were segregating both blue and white parental alleles in the blue-flowered pool, indicating that none of these comprises the causative locus. The same pattern was observed for four of the six transcription factors (the R2R3 MYB **AN2**, the bHLH **jaf13**, the WDR **AN11**, and the R3MYB **MYBx**). The **AN1** bHLH ortholog presented nearly undetectable floral expression (Supporting Information Table S3), and thus no SNPs could be confidently called across the pools. This result suggests that between **jaf13** and **AN1**, the former is likely the principal bHLH partner in the anthocyanin regulatory complex in **Iochroma** flowers. The final candidate locus, the R2R3 MYB repressor **MYB27** did present zero called variants in the blue-flowered pool, as would be predicted for a causal locus. However, this gene was not recovered in the transcriptome-wide search because of its low expression (Supporting Information Table S3) and accordingly low coverage. Such low floral expression is consistent with studies in **Petunia**, suggesting that its primary role is repression of anthocyanins in vegetative tissue (Albert **et al.**, 2011). Thus, the lack of called variants for **MYB27** in the blue pool is likely due to linkage with the causative locus. Indeed, additional bioinformatic analyses (see below) suggest that the **MYBL1** gene detected in the transcriptome-wide search is closely linked to the **MYB27** ortholog in **Iochroma**.

Novel R3 MYB belongs to a clade of R2R3 MYB repressors

The similarity of **Iochroma MYBL1** to an R2R3 MYB in tomato raises the possibility that this R3 MYB might be derived from a past duplication of an R2R3 MYB. Indeed, additional searches of nearby tomato scaffolds from chromosome 5 uncovered a highly similar sequence in an unannotated region that was 10kb from the first tomato R2R3 found with BLAST. This suggested that tomato might possess an ortholog of **MYBL1** that had been missed in annotation. Subsequent gene prediction analyses revealed a complete open reading frame with a highly conserved R3 MYB region in tomato as well as **S. pennellii**, **S. tuberosum**, and **C. annuum**, all members of the fleshy fruited subfamily, Solanoideae, which includes **Iochroma**. Although **P. hybrida** and **Coffea canephora** contained sufficiently large contigs (<15kb) around the conserved R2R3 MYBs, we did not recover orthologs of **MYBL1** from these taxa.

Phylogenetic analysis of **MYBL1** along with other R2R3 and R3 MYBs are consistent with a recent origin via duplication within the clade of R2R3 MYB repressors. The **MYBL1** sequences from **Iochroma**, **Capsicum** and **Solanum** form a well-supported clade (100% posterior probability, Fig. 2) that is closely related to a group of 'MYB3-like' Solanaceae R2R3 genes, including the tomato R2R3 originally uncovered in the BLAST search.

These R2R3 MYBs likely act as repressors due to the presence of C-terminal repression motifs (Jin **et al.**, 2000) (Fig. 3). Indeed, as mentioned above, one of these genes (**MYB27** in **Petunia hybrida**) has been well-characterized and actively represses anthocyanin biosynthesis (Albert **et al.**, 2011; Albert **et al.**, 2014). The presence of only a single **Petunia** sequence in this Solanaceae repressor clade suggests that the duplication which gave rise to **MYBL1** occurred after the split of Petunieae and the "x=12" clade containing Solanoideae (23-47 mya) (De-Silva **et al.**, 2017). These Solanaceae **MYBL1** and MYB3-like sequences are closely related to **AtMYBL2** and the **Arabidopsis** subgroup 4 MYBs (Fig. 2), which also function as transcription repressors (Dubos **et al.**, 2008; Dubos **et al.**, 2010).

Members of the **MYBL1** clade share several changes in their coding sequence that may relate to their functional evolution. First, these MYBs present large deletions in the R2 domain, comparable to **AtMYBL2** and other R3 MYBs (Fig. 3). These deletions have resulted in the loss of the helix-turn-helix motifs required for DNA-binding and functionality of the domain (Ogata **et al.**, 1994; Williams & Grotewold, 1997). Second, the **MYBL1** sequences have lost the C-terminal active repression motif $DLNx_2P$ present in the MYB3-like R2R3 sequences and in Petunia's **MYB27.** Nonetheless, they have gained a novel repression motif (LxLxL) near the end of the R3 domain (Fig. 3). Both of these repression motifs fall into the class of Ethylene-responsive element binding factor-associated Amphiphilic Repression or EAR motifs, which are found through plants (Kagale & Rozwadowski, 2011). Thus, it appears that the repressive activity of **MYBL1** was likely retained despite a roughly 40% reduction in the length of the protein. Neither the **MYBL1** clade nor **AtMYBL2** contain the WxM motif required for cell-to-cell movement (Kurata **et al.**, 2005), suggesting they do not diffuse like the smaller R3 MYBs of the CPC clade (Fig. 2).

Differential expression associated with color phenotypes

Comparison of gene expression levels across the backcross pools points to upregulation of **MYBL1** associated with loss of pigmentation, consistent with its putative role as a pigment repressor. In the backcross-to-blue, the F_1 phenotype individuals have roughly twice the **MYBL1** expression of the blue-flowered individuals (Fig. 4). Given that the trans-regulatory environment should be homogenized across these two pools, this difference would be best explained by a cis-regulatory change at **MYBL1**. Expression levels for **MYBL1** were slightly elevated in the white backcross individuals compared to those with the F1 phenotype, but the difference was not significant (Fig. 4). A marked difference in expression of **MYBL1** was also observed between the parents, although the **I. cyaneum** expression level overlapped with multiple backcross pools. These qPCR experiment produced very similar patterns, with the strongest differences between the two parents and between the two phenotypic pools in the backcross to blue (Supporting Information, Fig. S2). Given the relatively subtle differences in expression, additional experiments would be useful to better quantify allele-specific **MYBL1** expression across the individuals in these segregating populations.

By contrast with **MYBL1**, the other transcription factors showed no evidence of differential expression in association with the color phenotypes. The three transcription factors with quantifiable floral expression (**AN2**, **jaf13**, and **AN11**) showed statistically indistinguishable levels of expression across the backcross pools in both the transcriptomic analysis and the qPCR experiment (Fig. 4; Supporting Information, Fig. S2). The lack of variation in expression at these loci suggests that **MYBL1** does not regulate these transcription factors, and instead controls flower pigmentation through regulation of thestructural genes of the pathway.

Consistent with this hypothesis, we observed marked variation in the expression of structural genes. In particular, **F3'5'h**, **Dfr**, and **Ans** showed an almost complete loss of expression in both the white-flowered and F_1-phenotype pools, suggesting that their expression is strongly inhibited by the increase in **MYBL1** expression. The significant downregulation of these three loci was also recovered in the qPCR experiment (Supporting Information Fig. S2). Two of the early genes, **Chs** and **F3h**, also showed a trend toward lower expression with decreasing pigmentation, but these patterns were not consistently significant (Fig. 4, Supporting Information Fig. S2). By contrast with the other structural genes, **Chi** did not exhibit lowered expression in the less pigmented pools. This gene may be controlled by different transcriptional regulators as implicated by studies in several other systems (Dubos **et al.**, 2008; Yi **et al.**, 2010; Koehler **et al.**, 2012).

R3 MYB acts as a functional floral anthocyanin repressor in Nicotiana

Expression of **MYBL1** from **I. loxense** in **Nicotiana tabacum** (tobacco) resulted in a nearly complete loss of floral anthocyanins. All of the transformed lines presented the white-flowered phenotype, with 11 to 20 times lower absorption across them and 14 times lower absorbance on average (Fig. 5). **N. tabacum**, another member of the Solanaceae, is estimated to be 32 million years diverged from **Iochroma** (De-Silva **et al.**, 2017), indicating that the R3 MYB has the ability to repress anthocyanins in distantly related taxa. Also, the primary pigment in **N. tabacum** is cyanidin (McCarthy **et al.**, 2017), as opposed to delphinidin, which is in the blue-flowered **Iochroma**. Given that cyanidin, a derivative of dihydroquercetin, does not require F3'5'H activity, this result suggests that the level of downregulation at the other pathway genes (Fig. 4) is likely sufficient to abolish pigment production.

Discussion

Through a combination of crossing studies and bulk-segregant RNA-sequencing, our results indicate that a novel R3-MYB transcription factor, **MYBL1**, controls the intensity of floral pigmentation in **Iochroma**. The **MYBL1** allele from the white-flowered **I. loxense**acts in a nearly dominant fashion, with the F_1 hybrid between **I. loxense** and the blue-flowered **I. cyaneum** presenting only traces of anthocyanin pigmentation towards the mouth of the corolla. Patterns of gene expression suggest that this loss of pigmentation is due to downregulation of structural genes in the pathway, particularly the late anthocyanin-specific elements. Although we cannot exclude the possibility of causative coding sequence mutations at **MYBL1**, the elevated expression of this gene in F_1 phenotype and white-flowered pools suggests a **cis**-regulatory change as the underlying mechanism, analogous to the R3 MYB transcription factor controlling flower color in **Mimulus** (Yuan **et al.**, 2013). Under this scenario, the transition from the ancestral state of blue flowers to the derived state of white flowers along the **I. loxense** lineage could be explained by the increased expression of the **MYBL1** regulator leading to downregulation of anthocyanin pathway genes and suppression of floral pigment production.

While the structure of **MYBL1** resembles known R3MYB repressors, phylogenetic analyses demonstrate that **MYBL1** is evolutionarily distinct. Most of the previously described R3 MYB repressors, such as **CAPRICE** (**CPC**) in **Arabidopsis** and **MYBx** in petunia, fall into a single clade that is distantly related to **MYBL1** (Fig. 2). Similar to **MYBL1**, these genes lack an R2 domain and contribute to the regulation of anthocyanin production (Zhu **et al.**, 2009) along with other aspects of epidermal cell differentiation (Wada **et al.**, 1997; Schnittger **et al.**, 1999; Serna, 2008; Wang & Chen, 2014). While the closest relatives of the **CPC** clade of R3 MYBs remain unclear, the **MYBL1** clade is well resolved as a part of the subgroup 4 R2R3 repressors (Fig. 2). All of the

characterized subgroup 4 genes in **Arabidopsis** (**AtMYB4**, **AtMYB7**, and **AtMYB2**) regulate elements of the phenylpropanoid pathway, which includes anthocyanins (Jin **et al.**, 2000; Preston **et al.**, 2004; Fornale **et al.**, 2014). Petunia **MYB27** also belongs to this subgroup 4 clade and modulates anthocyanin production (Albert **et al.**, 2011). These patterns suggest that anthocyanin regulation is likely the ancestral state for the **MYBL1** clade as well as the other post-duplication copy, **MYB3-like**. Given the physical proximity of the **MYBL1** genes to the **MYB3like** R2R3 genes in the Solanoid genomes (ca. 10kb), we hypothesize that these R3 MYBs arose by tandem duplication followed by loss of the R2 domain.

The molecular evolution of the **MYBL1** clade within Solanaceae presents a striking parallel to the evolution of the **AtMYBL2** gene in **Arabidopsis**. Both of these single domain R3 MYB repressors arose from clades of R2R3 MYB repressors, and both show a dynamic history of loss and gain of repression motifs. **AtMYBL2** lost the EAR motif shared amongthe subgroup 4 R2R3 MYB repressors (Liu **et al.**, 2015) but gained a novel TLxLFR repression motif at the c-terminus (Matsui **et al.**, 2008). Similarly, **MYBL1** lost the EAR motif characteristic of the **MYB3-like** genes and Petunia **MYB27** (Albert **et al.**, 2011), but acquired a new EAR motif near the end of the R3 domain (Fig. 3). Like **AtMYBL2**, the **MYBL1** clade has retained the conserved bHLH binding motif in the R3 domain (Zimmermann **et al.**, 2004) and thus is likely to bind with the bHLH transcription factors and act as part of an MBW regulatory complex.

The presence of the EAR repression motif in **MYBL1** suggests that it functions as an active transcriptional repressor, although some degree of passive repression is possible. Given the structural similarity to **AtMYBL2**, we predict that **MYBL1** binds the bHLH factor and possibly also to the R2R3 partner (Albert **et al.**, 2014). The presence of the R2R3 MYB (**AN2** homolog), which normally acts as an activator, allows the complex to identify the transcriptional targets, while the R3 MYB induces epigenetic silencing via the EAR motif (Kagale & Rozwadowski, 2011). Thus, the replacement of one of the R2R3 MYB partners in the MBW complex with a R3 MYB could transform the complex from an activator of anthocyanin gene transcription to a repressor (Albert **et al.**, 2014) (Fig. 6). Depending on its abundance relative to the R2R3 activators, it is also possible that MYBL1 could passively repress anthocyanin gene expression by titrating bHLH factors, similarly to the CPC clade of R3 MYBs that lack the EAR active repression motifs (Mitsuda & Ohme-Takagi, 2009). Both mechanisms are consistent with the pattern seen in the backcrosses, where intermediate levels of **MYBL1** expression (as in the F_1) allow some pigment production (presumably because of some MBW complexes contain only R2R3 MYB activators). When the R3 MYB expression is high, as in the white-flowered pools and parent, most MBW complexes contain at least one R3 MYB, and thus function as repressors. Additional studies would be needed to confirm this model for **MYBL1** function.

The mode of action of **MYBL1** may have important consequences for understanding the evolutionary trajectory that led to the transition to white flowers in **I. loxense**. Judging from expression levels in the phenotypic pools (Fig. 4), a 30% increase in **MYBL1** expression might be sufficient to convert a blue flower into a nearly white flower. Even a small increase might be sufficient to produce a pale phenotype that could be targeted by selection, whether by biotic factors, like pollinators, or abiotic factors, such as climatic conditions (Rausher, 2008). We do not presently know the mutation(s) that caused the observed expression change in **MYBL1**, but if it were accomplished with a single mutation, the resulting allele could spread quickly through the population given its nearly dominant action (Haldane, 1924; Haldane, 1927). This stands in contrast to most of the previously identified genetic changes associated with transitions to white flowers, which involve loss-of-

function mutations in the R2R3 MYB activators (Quattrocchio **et al.**, 1999; Schwinn **et al.**, 2006; Hoballah **et al.**, 2007). These alleles are recessive and thus would experience a lower probability of fixation, a phenomenon known as "Haldane's Sieve" (Turner, 1981; Charlesworth, 1992).

Although loss-of-function R2R3 MYB mutations and gain-of-function R3 MYB mutations are expected to experience different rates of fixation, their ultimate effect on development is likely to be similar. Because the R3 MYB acts by partnering with the bHLH and using the R2R3 MYB to identify targets, the set of anthocyanin genes which experience downregulation are expected to be the same. In the case of **Iochroma** and other asterids, the MBW complex, whether acting as an activator or repressor, most strongly targets the late genes of the pathway. The isolation of effects to the late genes may be important in allowing the suppression of pigment production but maintenance of other pathway products, like flavones and flavonols (Streisfeld & Rausher, 2011; Wessinger & Rausher, 2012). Similarly, the apparent bias towards fixation of regulatory mutations as opposed to coding mutations in structural genes may reflect selection to maintain the intact pathway for pigment production in other tissues and developmental stages (Wessinger & Rausher, 2012; Ho & Smith, 2016). Overall, this work, together with studies in other flowering plants suggest that convergence at the level of individual genes and mutations may not be expected when the underlying mechanisms present multiple, developmentally similar, pathways to the same phenotype.

Acknowledgements

This research was supported by funding from the NIH (F32GM080082) and NSF (DEB-1355518 to SDS and MCB-1412738 to BJSCO) as well as a Layman Award from the University of Nebraska-Lincoln. The authors thank the Ministerio del Ambiente de Ecuador for collecting permits (N °013-IC-FLO-DBAP-MA), M. D. Rausher for early input on the project and three anonymous reviewers for feedback.

Author contributions

S.D.S. and D.J.G. designed and carried out the experiments and analyzed the data; B.J.S.C.O. collaborated on the bioinformatic analysis; T.E.C. contributed materials and expertise to the functional test in **Nicotiana**. D.J.G. and S.D.S. wrote the paper, and all authors contributed to revision of the manuscript.

Figure Legends

Fig. 1 Phenotypic variation in parents and backcrosses. (a) Flowers and corolla mouths of parental species (**I. cyaneum and I. loxense**), their F_1, and example phenotypes from the backcrosses. (b) Boxplots of anthocyanin content in corolla tissue, measured as absorbance at 530nm/g fresh tissue. Order of the phenotypes along the x-axis follows (a). Boxes encompass the upper and lower quartiles and the horizontal line represents the median.

Fig. 2 Majority rule consensus tree from Bayesian analysis of MYB sequences. Posterior probabilities are given above branches. Sequences are prefixed by TAIR or GenBank numbers. Known repressors are highlighted. Taxon names are abbreviated as follows: Slyc, **Solanum lycopersicum**; Stub, potato; Spen, **S. pennellii**; Cann, **Capsicum annuum**; Ph, **Petunia hybrida**; Ccan, (**Coffea canephora**); Ml, **Mimulus lewisii**; Am, **Antirrhinum**

majus; and At, **Arabidopsis thaliana**. The suffix for At sequences includes the subgroup number when available (Dubos **et al.**, 2010).

Fig. 3 Comparison of MYBL1 and closely related sequences. The DNA-binding R2 and R3 domains are indicated; each contains three α-helices (gray boxes) with conserved tryptophan residues (*). The R3 domain also includes a bHLH binding motif. Locations from Stracke et al. (2014) and Du et al. (2015). Repression motifs include EAR motifs and the TLxLFR motif. These often fall within the C-terminal motifs (C1, C2, C3, C4), which are conserved across the subgroup 4 repressors (Kranz **et al.**,1998; Legay **et al.**, 2007). Note that these sequence motifs were identified in **Arabidopsis** and differ slightly in Solanaceae.

Fig. 4 Variation in pathway gene expression across phenotypic pools. A simplified pathway is depicted on the left comprising the enzymes chalcone synthase (CHS), chalcone isomerase (CHI), flavonone 3-hydroxylase (F3H), flavonoid 3 ′-hydroxylase (F3’H), flavonoid 3 ′,5 ′ -hydroxylase (F3’5’H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS). Branches of the pathway leading to pelargonidin and cyanidin pigments are not shown as **I. cyaneum** produces only delphinidin. Note that in **Iochroma**, F3’5’H has little 3’ activity and thus appears not to act on dihydrokaempferol (Smith & Rausher, 2011). Elements of the pathway are regulated by bHLH, WDR, and R2R3 and R3 MYB proteins. Expression levels are given for each of these loci in fragments per kilobase per million (FPKM). Error bars denote 95% confidence intervals, and letters indicate significantly different values.

Fig. 5 Effects of **MYBL1** expression in **N. tabacum**. Anthocyanin content is measured as absorbance at 530nm/g fresh tissue. Error bars show +/- 1 standard deviation. Representative flowers are shown for each treatment.

Fig. 6 Model for R3 MYB repression of anthocyanin production. (a) Activation of anthocyanin biosynthetic genes (ABG) by the MBW complex based on the Albert et al. (2014) model. The MYB and bHLH proteins are thought to be present as dimers within the complex, with MYBs being bridged through their binding to the bHLH factors. (b) The R3 MYB (dark circle) could block anthocyanin production by titrating bHLH factors (and their bound WD40 partners) as passive repressors (left) or by partnering with R2R3 MYB activators as part of MBW complexes and triggering active (transcriptional) repression (right).

Supporting Information

Table S1 Primer pairs for qPCR.

Table S2 Thirty-five candidate genes from first bioinformatics filter.

Table S3 SNP frequencies and expression levels for anthocyanin pathway genes.

Fig. S1 Principal components analysis of floral color spectra.

Fig. S2 Differences in gene expression based on qualitative real-time PCR.

References

Albert NW, Davies KM, Lewis DH, Zhang HB, 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**: 771-784.

Berardi AE, Hildreth SB, Helm RF, Winkel BSJ, Smith SD. 2016. Evolutionary correlations in flavonoid production across flowers and leaves in the Iochrominae (Solanaceae). Phytochemistry **130**: 119-127.

Carrington JC, Freed DD. 1990. Cap-Independent enhancement of translation by a plant potyvirus 5' nontranslated region. Journal of Virology **64**: 1590-1597.

Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. American Naturalist **140**: 126-148.

Clemente T. 2006. Nicotiana (Nicotiana tobaccum, Nicotiana benthamiana). High-Throughput Next Generation Sequencing: Methods and Application **343**: 143-154.

Coburn RA, Griffin RH, Smith SD. 2015. Genetic basis for a rare floral mutant in an Andean species of Solanaceae. American Journal of Botany **102**: 264-272.

Davies KM, Albert NW, Schwinn KE. 2012. From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigmentation patterning. Functional Plant Biology **39**: 619-638.

De-Silva DL, Mota LL, Chazot N, Mallarino R, Silva-Brandao KL, Pinerez LMG, Freitas AVL, Lamas G, Joron M, Mallet J, et al. 2017. North Andean origin and diversification of the largest ithomiine butterfly genus. Scientific Reports **7**: 45966.

de Vlaming P, Gerats AGM, Wiering H, Wijsman HJW. 1984. Petunia hybrida: A short description of the action of 91 genes, their origin, and their map location. Plant Molecular Biology Reporter **2**: 21-42.

Du H, Liang Z, Zhao S, Nan MG, Tran LSP, Lu K, Huang YB, Li JN. 2015. The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: New insights into subfamily classification and expansion. Scientific Reports **5**: 11037.

Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E, Debeaujon I, Routaboul JM, Alboresi A, Weisshaar B, Lepiniec L. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in Arabidopsis thaliana. Plant Journal **55**: 940-953.

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.

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

Fornale S, Lopez E, Salazar-Henao JE, Fernandez-Nohales P, Rigau J, Caparros-Ruiz D. 2014. AtMYB7, a new player in the regulation of UV-sunscreens in Arabidopsis thaliana. Plant and Cell Physiology **55**: 507-516.

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

Grotewold E. 2006. The genetics and biochemistry of floral pigments. Annual Review of Plant Biology **57**: 761-780.

Hajdukiewicz P, Svab Z, Maliga P. 1994. The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. Plant Molecular Biology **25**: 989-994.

Haldane JBS. 1924. A mathematical theory of natural and artificial selection - Part II - The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of mendelian populations, and on natural selection. Proceedings of the Cambridge Philosophical Society-Biological Sciences **1**: 158-163.

Haldane JBS. 1927. A mathematical theory of natural and artificial selection, Part V: Selection and mutation. Proceedings of the Cambridge Philosophical Society **23**: 838-844.

Ho WW, Smith SD. 2016. Molecular evolution of anthocyanin pigmentation genes following losses of flower color. BMC Evolutionary Biology **16**: 98.

Hoballah ME, Gubitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo A, Arnold M, Kuhlemeier C. 2007. Single gene-mediated shift in pollinator attraction in Petunia. Plant Cell **19**: 779-790.

Jin HL, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO Journal **19**: 6150-6161.

Kagale S, Rozwadowski K. 2011. EAR motif-mediated transcriptional repression in plants: Anunderlying mechanism for epigenetic regulation of gene expression. Epigenetics **6**: 141-146.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology **14**: R36.

Koehler G, Wilson RC, Goodpaster JV, Sonsteby A, Lai X, Witzmann FA, You JS, Rohloff J, Randall SK, Alsheikh M. 2012. Proteomic study of low-temperature responses in strawberry cultivars (Fragaria x ananassa) that differ in cold tolerance. Plant Physiology **159**: 1787-1805.

Koes R, Verweij W, Quattrocchio F. 2005. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends in Plant Science **10**: 236-242.

Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C, et al. 1998. Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. Plant Journal **16**: 263-276.

Kurata T, Ishida T, Kawabata-Awai C, Noguchi M, Hattori S, Sano R, Nagasaka R, Tominaga R, Koshino-Kimura Y, Kato T, et al. 2005. Cell-to-cell movement of the CAPRICE protein in Arabidopsis root epidermal cell differentiation. Development **132**: 5387-5398.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10**: R25.

Legay S, Lacombe E, Goicoechea M, Briere C, Seguin A, Mackay J, Grima-Pettenati J. 2007. Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. Plant Science **173**: 542-549.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078-2079.

Liu JY, Osbourn A, Ma PD. 2015. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Molecular Plant **8**: 689-708.

Matsui K, Umemura Y, Ohme-Takagi M. 2008. AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. Plant Journal **55**: 954-967.

McCarthy EW, Berardi AE, Smith SD, Litt A. 2017. Related allopolyploids display distinct floral pigment profiles and transgressive pigments. American Journal of Botany **104**: 92-101.

Mitsuda N, Ohme-Takagi M. 2009. Functional analysis of transcription factors in Arabidopsis. Plant and Cell Physiology **50**: 1232-1248.

Ng J, Smith SD. 2016. Widespread flower color convergence in Solanaceae via alternate biochemical pathways. New Phytologist **209**: 407-417.

Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y. 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell **79**: 639-648.

Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW. 2004. AtMYB32 is required for normal pollen development in Arabidopsis thaliana. Plant Journal **40**: 979-995.

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**: 1433-1444.

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

Rausher MD. 2008. Evolutionary transitions in floral color. International Journal of Plant Sciences **169**: 7-21.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**: 1572-1574.

Schnittger A, Folkers U, Schwab B, Jurgens G, Hulskamp M. 1999. Generation of a spacing pattern: The role of TRIPTYCHON in trichome patterning in Arabidopsis. Plant Cell **11**: 1105-1116.

Schwinn K, Venail J, Shang YJ,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.

Serna L. 2008. CAPRICE positively regulates stomatal formation in the Arabidopsis hypocotyl. Plant Signal Behav **3**: 1077-1082.

Smith SD, Baum DA. 2006. Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). American Journal of Botany **93**: 1140-1153.

Smith SD, Baum DA. 2007. Systematics of Iochrominae (Solanaceae): Patterns in floral diversity and interspecific crossability. Acta Horticulturae **745**: 241-254.

Smith SD, Goldberg EE. 2015. Tempo and mode of flower color evolution. American Journal of Botany **102**: 1-12.

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

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

Stracke R, Holtgrawe D, Schneider J, Pucker B, Sorensen TR, Weisshaar B. 2014. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (Beta vulgaris). BMC Plant Biology **14**: 249.

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, Rausher MD. 2011. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. Evolution **65**: 629-642.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology **28**: 511-515.

Turner JRG. 1981. Adaptation and evolution in Heliconius: a defense of Neodarwinism. Annual Review of Ecology and Systematics **12**: 99-121.

van Houwelingen A, Souer E, Spelt K, Kloos D, Mol J, Koes R. 1998. Analysis of flower pigmentation mutants generated by random transposon mutagenesis in Petunia hybrida. Plant Journal **13**: 39-50.

Wada T, Tachibana T, Shimura Y,Okada K. 1997. Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. Science **277**: 1113-1116.

Wang SC, Chen JG. 2014. Regulation of cell fate determination by single-repeat R3 MYB transcription factors in Arabidopsis. Frontiers in Plant Science **5**:133.

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

Williams CE, Grotewold E. 1997. Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities. Journal of Biological Chemistry **272**: 563-571.

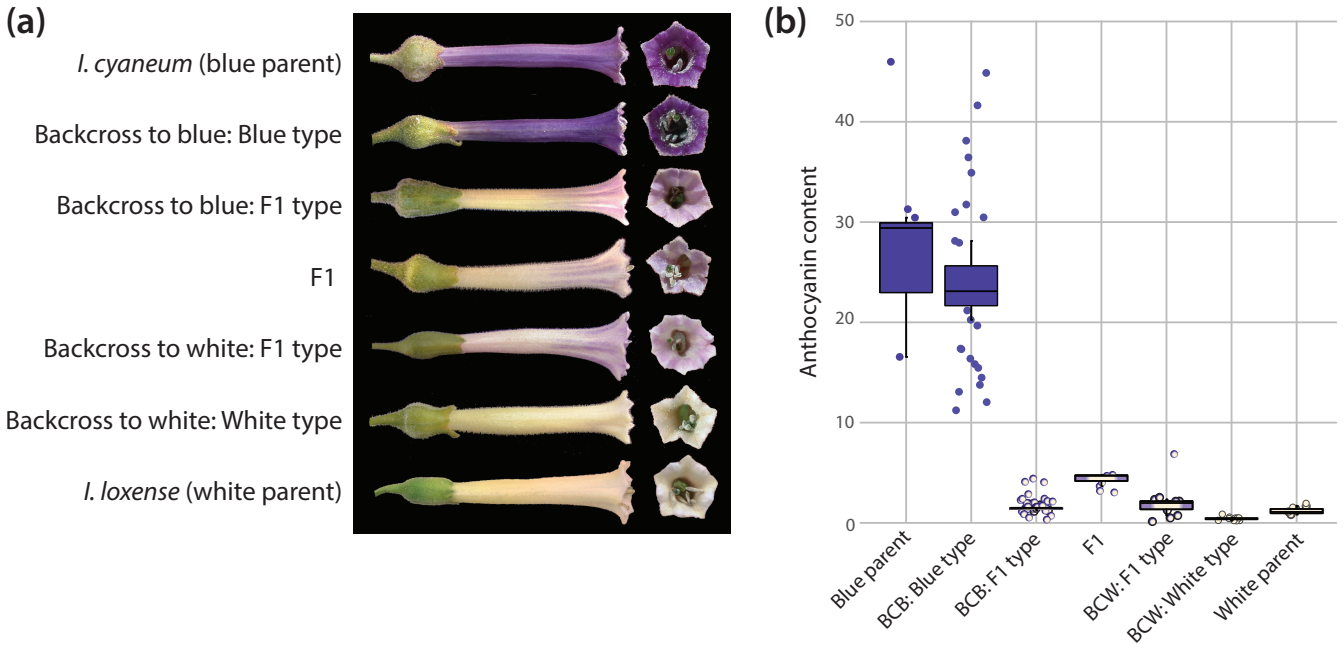
Winkel-Shirley B. 2001. Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology and biotechnology. Plant Physiology **126**: 485-492.

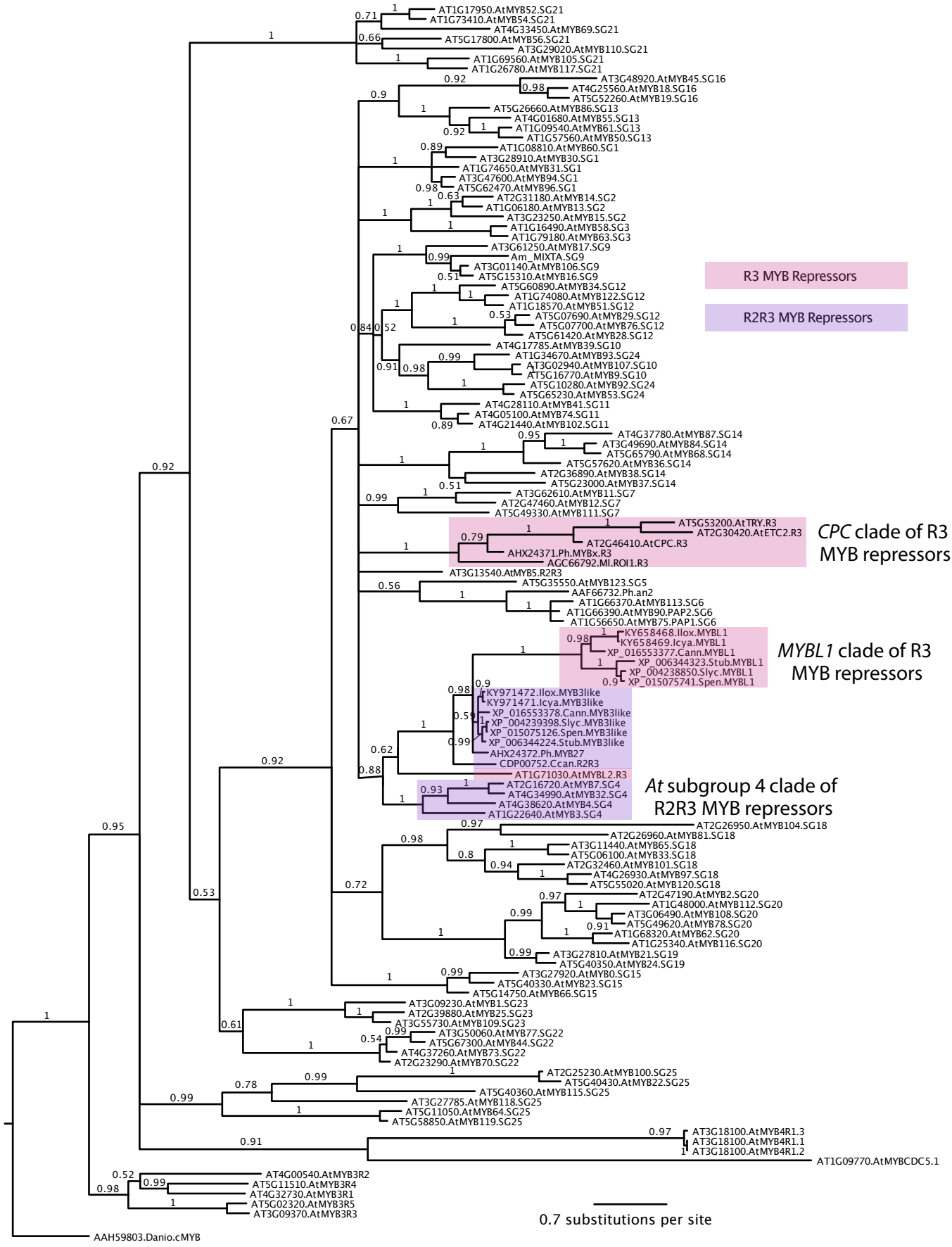
Yi JX, Derynck MR, Li XY, Telmer P, Marsolais F, Dhaubhadel S. 2010. A single-repeat MYB transcription factor, GmMYB176, regulates CHS8 gene expression and affects isoflavonoid biosynthesis in soybean. Plant Journal **62**: 1019-1034.

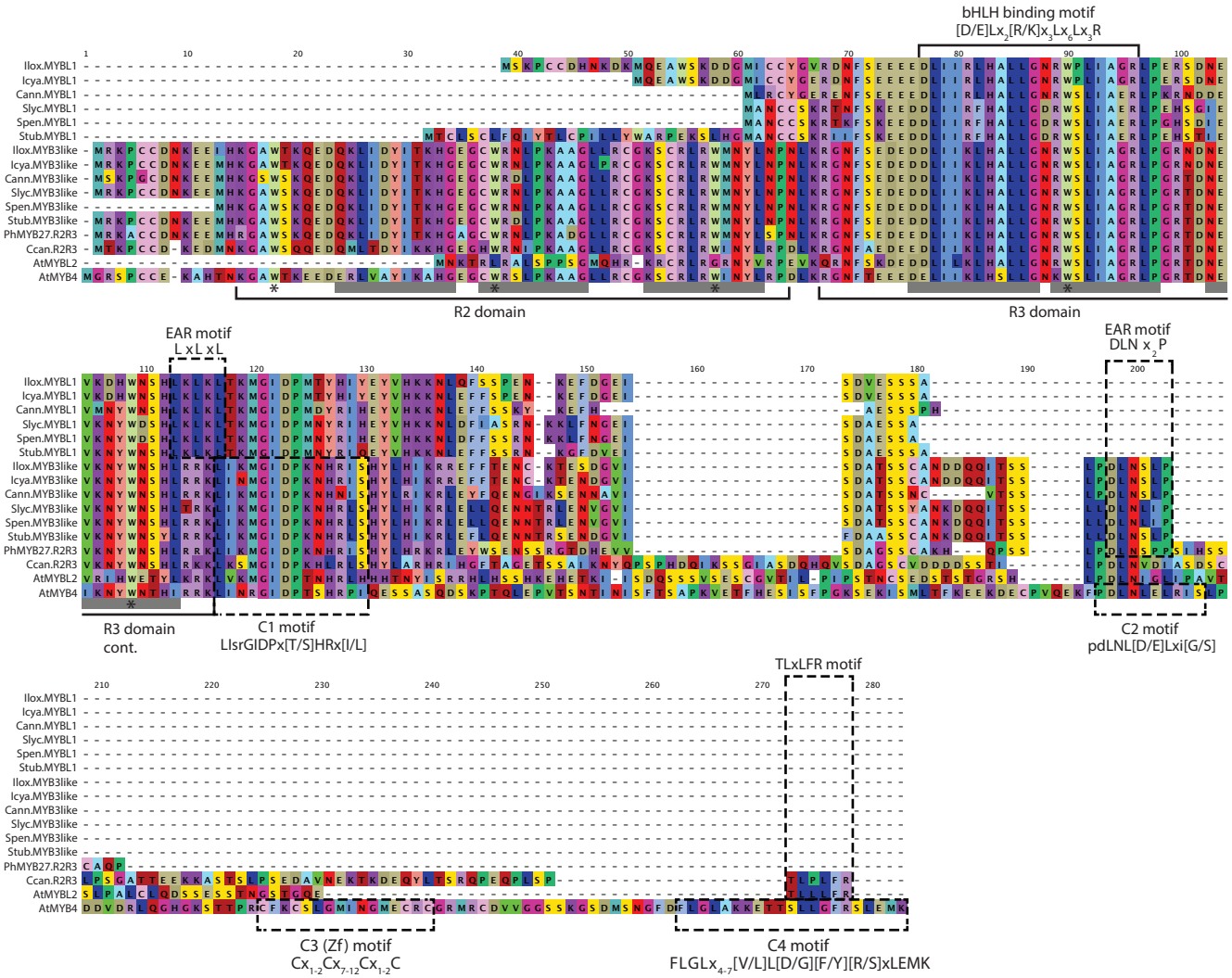
Yuan YW, Sagawa JM, Young RC, Christensen BJ, Bradshaw HD. 2013. Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of Mimulus. Genetics **194**: 255-263.

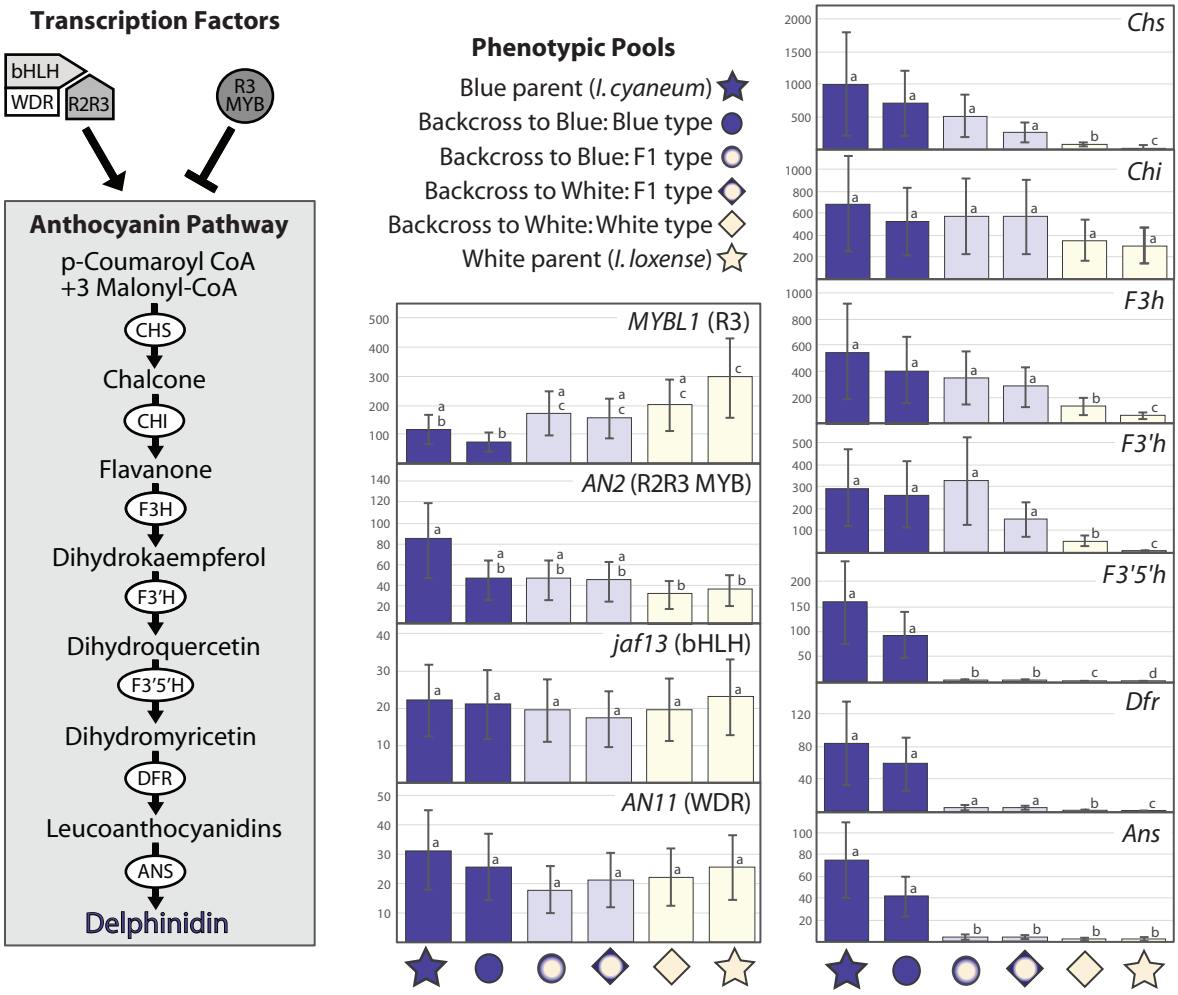
Zhu HF, Fitzsimmons K, Khandelwal A, Kranz RG. 2009. CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. Molecular Plant **2**: 790-802.

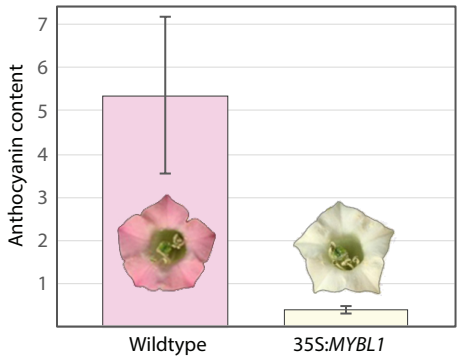
Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. 2004. Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. Plant Journal **40**: 22-34.

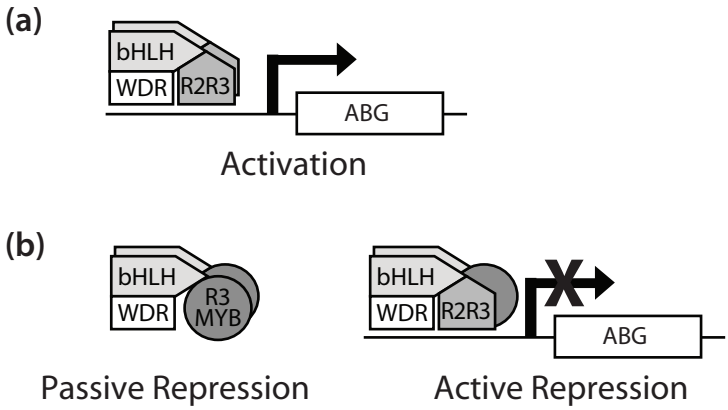

(a)
I. cyaneum (blue parent)
Backcross to blue: Blue type
Backcross to blue: F1 type
F1
Backcross to white: F1 type
Backcross to white: White type
I. loxense (white parent)
(b)
Anthocyanin content
50
40
30
20
10
0
Blue parent
BCB: Blue type
BCB: F1 type
F1
BCW: F1 type
BCW: White type
White parent


AT1G17950.AtMYB52.SG21
AT1G73410.AtMYB54.SG21
AT4G33450.AtMYB69.SG21
AT5G17800.AtMYB56.SG21
AT3G29020.AtMYB110.SG21
AT1G69560.AtMYB105.SG21
AT1G26780.AtMYB117.SG21
AT3G48920.AtMYB45.SG16
AT4G25560.AtMYB18.SG16
AT5G52260.AtMYB19.SG16
AT5G26660.AtMYB86.SG13
AT4G01680.AtMYB55.SG13
AT1G09540.AtMYB61.SG13
AT1G57560.AtMYB50.SG13
AT1G08810.AtMYB60.SG1
AT3G28910.AtMYB30.SG1
AT1G74650.AtMYB31.SG1
AT3G47600.AtMYB94.SG1
AT5G62470.AtMYB96.SG1
AT2G31180.AtMYB14.SG2
AT1G06180.AtMYB13.SG2
AT3G23250.AtMYB15.SG2
AT1G16490.AtMYB58.SG3
AT1G79180.AtMYB63.SG3
AT3G61250.AtMYB17.SG9
Am_MIXTA.SG9
AT3G01140.AtMYB106.SG9
AT5G15310.AtMYB16.SG9
AT5G60890.AtMYB34.SG12
AT1G74080.AtMYB122.SG12
AT1G18570.AtMYB51.SG12
AT5G07690.AtMYB29.SG12
AT5G07700.AtMYB76.SG12
AT5G61420.AtMYB28.SG12
AT4G17785.AtMYB39.SG10
AT1G34670.AtMYB93.SG24
AT3G02940.AtMYB107.SG10
AT5G16770.AtMYB9.SG10
AT5G10280.AtMYB92.SG24
AT5G65230.AtMYB53.SG24
AT4G28110.AtMYB41.SG11
AT4G05100.AtMYB74.SG11
AT4G21440.AtMYB102.SG11
AT4G37780.AtMYB87.SG14
AT3G49690.AtMYB84.SG14
AT5G65790.AtMYB68.SG14
AT5G57620.AtMYB36.SG14
AT2G36890.AtMYB38.SG14
AT5G23000.AtMYB37.SG14
AT3G62610.AtMYB11.SG7
AT2G47460.AtMYB12.SG7
AT5G49330.AtMYB111.SG7
AT5G53200.AtTRY.R3
AT2G30420.AtETC2.R3
AT2G46410.AtCPC.R3
AHX24371.Ph.MYBx.R3
AGC66792.Ml.ROI1.R3
AT3G13540.AtMYB5.R2R3
AT5G35550.AtMYB123.SG5
AAF66732.Ph.an2
AT1G66370.AtMYB113.SG6
AT1G66390.AtMYB90.PAP2.SG6
AT1G56650.AtMYB75.PAP1.SG6
KY658468.Ilox.MYBL1
KY658469.Icya.MYBL1
XP_016553377.Cann.MYBL1
XP_006344323.Stub.MYBL1
XP_004238850.Slyc.MYBL1
XP_015075741.Spen.MYBL1
KY971472.Ilox.MYB3like
KY971471.Icya.MYB3like
XP_016553378.Cann.MYB3like
XP_004239398.Slyc.MYB3like
XP_015075126.Spen.MYB3like
XP_006344224.Stub.MYB3like
AHX24372.Ph.MYB27
CDP00752.Ccan.R2R3
AT1G71030.AtMYBL2.R3
AT2G16720.AtMYB7.SG4
AT4G34990.AtMYB32.SG4
AT4G38620.AtMYB4.SG4
AT1G22640.AtMYB3.SG4
AT2G26950.AtMYB104.SG18
AT2G26960.AtMYB81.SG18
AT3G11440.AtMYB65.SG18
AT5G06100.AtMYB33.SG18
AT2G32460.AtMYB101.SG18
AT4G26930.AtMYB97.SG18
AT5G55020.AtMYB120.SG18
AT2G47190.AtMYB2.SG20
AT1G48000.AtMYB112.SG20
AT3G06490.AtMYB108.SG20
AT5G49620.AtMYB78.SG20
AT1G68320.AtMYB62.SG20
AT1G25340.AtMYB116.SG20
AT3G27810.AtMYB21.SG19
AT5G40350.AtMYB24.SG19
AT3G27920.AtMYB0.SG15
AT5G40330.AtMYB23.SG15
AT5G14750.AtMYB66.SG15
AT3G09230.AtMYB1.SG23
AT2G39880.AtMYB25.SG23
AT3G55730.AtMYB109.SG23
AT3G50060.AtMYB77.SG22
AT5G67300.AtMYB44.SG22
AT4G37260.AtMYB73.SG22
AT2G23290.AtMYB70.SG22
AT2G25230.AtMYB100.SG25
AT5G40430.AtMYB22.SG25
AT5G40360.AtMYB115.SG25
AT3G27785.AtMYB118.SG25
AT5G11050.AtMYB64.SG25
AT5G58850.AtMYB119.SG25
AT3G18100.AtMYB4R1.3
AT3G18100.AtMYB4R1.1
AT3G18100.AtMYB4R1.2
AT1G09770.AtMYBCDC5.1
AT4G00540.AtMYB3R2
AT5G11510.AtMYB3R4
AT4G32730.AtMYB3R1
AT5G02320.AtMYB3R5
AT3G09370.AtMYB3R3
AAH59803.Danio.cMYB
R3 MYB Repressors
R2R3 MYB Repressors
CPC clade of R3 MYB repressors
MYBL1 clade of R3 MYB repressors
At subgroup 4 clade of R2R3 MYB repressors
0.7 substitutions per site


bHLH binding motif
[D/E]Lx2[R/K]x3Lx6Lx3R
Ilox.MYBL1
Icya.MYBL1
Cann.MYBL1
Slyc.MYBL1
Spen.MYBL1
Stub.MYBL1
Ilox.MYB3like
Icya.MYB3like
Cann.MYB3like
Slyc.MYB3like
Spen.MYB3like
Stub.MYB3like
PhMYB27.R2R3
Ccan.R2R3
AtMYBL2
AtMYB4
R2 domain
R3 domain
EAR motif
LxLxL
EAR motif
DLN x2 P
R3 domain
cont.
C1 motif
LIsrGIDPx[T/S]HRx[I/L]
C2 motif
pdLNL[D/E]Lxi[G/S]
TLxLFR motif
C3 (Zf) motif
Cx1-2Cx7-12Cx1-2C
C4 motif
FLGLx4-7[V/L]L[D/G][F/Y][R/S]xLEMK


Transcription Factors
bHLH
WDR
R2R3
R3
MYB
Anthocyanin Pathway
p-Coumaroyl CoA
+3 Malonyl-CoA
CHS
Chalcone
CHI
Flavanone
F3H
Dihydrokaempferol
F3'H
Dihydroquercetin
F3'5'H
Dihydromyricetin
DFR
Leucoanthocyanidins
ANS
Delphinidin
Phenotypic Pools
Blue parent (I. cyaneum)
Backcross to Blue: Blue type
Backcross to Blue: F1 type
Backcross to White: F1 type
Backcross to White: White type
White parent (I. loxense)
MYBL1 (R3)
AN2 (R2R3 MYB)
jaf13 (bHLH)
AN11 (WDR)
Chs
Chi
F3h
F3'h
F3'5'h
Dfr
Ans


Anthocyanin content
7
6
5
4
3
2
1
Wildtype
35S:MYBL1


(a)
bHLH
WDR
R2R3
ABG
Activation
(b)
bHLH
WDR
R3
MYB
bHLH
WDR
R2R3
ABG
Passive Repression
Active Repression